

Draft #4



United States Department of Agriculture

**Geospatial Data
Acquisition, Integration, and Delivery
National Implementation Strategy Plan**

September 22, 1999

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	4
2	INTRODUCTION.....	6
2.1	BACKGROUND	6
2.2	PURPOSE.....	7
2.3	SCOPE.....	7
2.3.1	Geospatial Data	8
2.3.2	Acquisition	8
2.3.3	Integration.....	8
2.3.4	Delivery	8
2.4	ACKNOWLEDGEMENTS.....	8
3	ACRONYMS	10
4	ACTIVITIES	13
4.1	OVERVIEW	13
4.2	STUDY OF AS-IS PROCESSES	13
4.2.1	Hardcopy Photography	14
4.2.2	Common Land Units	14
4.2.3	Soils.....	15
4.2.4	Legacy Systems	15
4.3	BEST PRACTICES AND OUTREACH EFFORTS.....	16
4.3.1	Corps of Engineers Best Practice Session.....	16
4.3.2	Beltsville Office Best Practice Forum.....	18
4.4	PILOT SITE SUPPORT ACTIVITIES.....	20
4.4.1	Data Themes	21
4.4.2	Critical Data Themes	22
4.4.3	Non-Critical Data Themes	22
4.4.4	Determining Future Data Themes.....	23
4.4.5	Geospatial Data Directory Structure and File Naming Conventions	23
4.4.6	Pilot Data Refresh.....	24
4.5	REVIEW OF THE PILOT DATA WITH SERVICE CENTER PERSONNEL.....	24
4.6	LESSONS LEARNED FROM THE PILOT SUPPORT PROCESS	27
5	GEOSPATIAL DATA STATUS	30
5.1	STATUS MAPS.....	30
5.2	THEME SUMMARIES.....	40
5.3	BUSINESS USES	42
5.4	PRODUCTION COSTS.....	44
5.5	PRODUCTION SCHEDULE.....	47
5.5.1	Critical Themes	47
5.5.2	Non-Critical Themes	49
6	AID ASSESSMENT	51
6.1	ACQUISITION.....	51
6.1.1	Problem Statements	51
6.1.2	Case Studies	52
6.1.3	Lessons Learned.....	54
6.2	INTEGRATION.....	56
6.2.1	Problem Statement	56
6.2.2	Case Studies	59
6.2.3	Lessons Learned.....	61
6.3	DELIVERY.....	62

6.3.1	Problem Statements	62
6.3.2	Case Studies	63
6.3.3	Lessons Learned.....	64
7	RECOMMENDATIONS	66
8	RISK ASSESSMENT AND MANAGEMENT IMPLICATIONS	70
APPENDIX A: Integration Examples		
APPENDIX B: Geospatial Data Acquisition, Integration and Delivery Project Plan		
APPENDIX C: BPR Pilot Sites Non-Critical Data Theme List		
APPENDIX D: USDA Service Center Initiative Geospatial Data Directory Structure and File		
Naming Convention - v5.0		
APPENDIX E: Beltsville Office Best Practice Forum Minutes USDA Service Center Initiative GIS		
Best Management Practices Forum		

1 Executive Summary

The Data Acquisition, Integration and Delivery Team was formed in May 1998 to address data issues impacting critical business areas of the three primary Service Center agencies. The Team was charged with establishing processes and policies for the acquisition, integration and delivery of the best available and ready to use geospatial data. The Team addressed the primary business needs of Natural Resources Conservation Service (NRCS), Field Services Agency (FSA), and Rural Development (RD), through the use of studies, evaluations of data management producers, development of prototypes, and implementing geospatial data through pilot sites.

Team expertise was derived from FSA and NRCS staff representing national and state agency leadership roles. Core team members were highly experienced technical staff familiar with GIS, data management, remote sensing, data delivery, spatial analysis, and relational database technology.

The recommendations put forth in this document are believed to offer the highest degree of efficiency for data delivery to field locations using current technology. These recommendations are intended to support the use of Service Center business applications while also helping to support the long-term development of a data infrastructure to serve USDA needs. The recommendations fell into nine categories and are summarized below; each is addressed in greater detail within the body of the document.

- 1) Fund APFO and NCGC at \$2 million, consistent with the President's FY2000 budget to acquire, integrate and deliver Mosaicked Digital Orthophoto Quadrangles (MDOQ), Common Land Unit (CLU), and soil data (SSURGO) for approximately 350 counties. Additional funds in future years will be based on production experience and process efficiencies.
- 2) Establish a National GIS Coordination Team to implement the recommendations in this strategy and guide GIS implementation. This Team should be composed of appropriate Business Process Re-engineering (BPR) team representatives, data stewards, existing acquisition, integration, and delivery (AID) representatives from APFO and NCGC, national headquarters GIS leads, state GIS specialists, and a representative of the Support Services Bureau (SSB) Data Management Team. The Deputy Chief of Soil Survey and Resource Assessment of NRCS and the Deputy Administrator for Farm Programs for FSA should be executive sponsors.
- 3) Continue to fund research and development efforts in data acquisition, integration, and delivery.
- 4) Deliver geospatial data to Service Center offices as a turnkey process that minimizes the data management requirements at the local level.
- 5) Monitor geospatial data requirements as business areas identify the data sets required to support their business needs.
- 6) Enforce adherence to geospatial standards, as recognized by the Data Management Team, for development and documentation of geospatial data sets.

- 7) Educate USDA users about geospatial data and their use limitations due to scale, resolution, map projection, and geometric and temporal mismatches. The GIS Training Team's materials should thoroughly address these issues.
- 8) Clearly define data access and data reproduction privileges for all potential users of the geospatial data.
- 9) Monitor acquisition, integration and delivery processes and respond to technological advances and availability of resources.
- 10) Involve state GIS staff in the Service Center Initiative (SCI) program, particularly with AID activities. State GIS staff should be provided the tools needed to assist APFO and NCGC with delivery of nationally-consistent, integrated geospatial data to Service Centers, as well as standards and specifications for supplementing high-resolution state and local data.

Embodied in recommendation #2 is the coordinated cross-agency prioritization of data development to support the SCI. Coordination would expedite the availability of the four critical data themes needed for business applications and may assist agencies in developing local partnerships and cooperative agreements. Presently, data sets are prioritized for digital development based on individual agency criteria.

Adoption of new technology carries risks for any organization. However in USDA it is compounded by many variables associated with funding, staff, and procurement. To be successful, adequate funding, in a timely fashion is needed for staffing, software, support, hardware and system architecture. It is imperative that all of these efforts are recognized in both the short-term and the long-term strategic planning initiatives in order for USDA to fully realize the benefits of the re-engineering process.

In the absence of financial support for any key component, USDA will fail in its efforts to provide integrated, consistent, easy-access, reliable information to staff, partners and cooperators. Multiple and inconsistent data sets will continue and data management inefficiencies will rob field staff of valuable time, which could better be used assisting landowners and fulfilling strategic initiatives.

USDA agencies have gained significant respect in recent years for innovation and creativity in the development of critical geo-spatial data themes. This is a direct result of historical investments in digital data development and staff training. To take full advantage of the existing and future data and allow customers to reap the benefit, this and other initiatives need to be adequately supported and funded by USDA. Data development and integration require long term commitments to ensure successful re-tooling of USDA business process.

2 INTRODUCTION

As we approach the millennium, how Service Centers do their daily business of conservation planning, lending, risk management, and community development is changing dramatically. The use of computerized spatial information rather than paper map products in daily business practices is the most fundamental shift our field staffs will experience in the next few years. Meeting the challenges of providing geospatial data in support of the Service Centers is a major undertaking.

To meet these challenges, the Geospatial Data Acquisition, Integration, and Delivery team was formed. Our guiding documents and driving forces included the *USDA Service Center Geographic Information System (GIS) Strategy* document, April 17, 1998; *USDA GIS Strategy for Service Centers*, 1998; *Data Rich and Information Poor*, November, 1995; *Future Directions: A Vision of Information Technology for Field Conservationist*, August, 1997; Service Center Implementation Team (SCIT) strategic plans; and Agency strategic plans (FSA, NRCS, RD).

The approaches of the Data AID team include a pilot project to acquire, integrate, and deliver geospatial data, consult with federal and state agencies that have been involved in similar projects, and to solicit information from Service Center staffs. Knowledge gained from these is reflected in this document.

This document, along with the *USDA Service Center Geographic Information System (GIS) Strategy* document dated April 17, 1998, and the *Service Center Data Implementation Plan*, currently in development, will guide the implementation of enterprise and Service Center GIS in the USDA.

2.1 Background

The Geospatial Data AID National Implementation Strategy Plan represents the outcomes and recommendations of the Data AID BPR (Business Process Re-engineering) Project Team. The team convened its first meeting at APFO (Aerial Photography Field Office) in Salt Lake City, Utah, in May 1998. At that meeting, the following team charter was established:

The AID project team will define and establish the processes, policies and provide oversight for the enterprise-wide acquisition, integration and delivery of "best available" ready to use geospatial data themes. The AID Team will address the business needs of Natural Resources Conservation Service, Field Services Agency, and Rural Development, through the use of studies, testing data management producers, and development of prototypes in defining the business case and implementation of the geospatial data.

In addition, the Data AID project expects the following non-quantifiable benefits to be derived from a re-engineered business process and application of automation technology:

- Enable "one-stop shopping".
- Present a detailed, consistent, and high quality representation of customer land, resulting in reduced costs and improved customer service.
- Enable easy Service Center and customer access to up-to-date, high-quality data.
- Allow Service Center agencies to jointly maintain a Common Land Unit (CLU) layer and provide consistent information to the customer.
- Facilitate information exchange between Service Centers and customers.
- Enable other federal (e.g., Forest Service, Bureau of Land Management, National Park Service, Environmental Protection Agency, etc.), state, and local agencies to use non-sensitive digital data developed and maintained by Service Center agencies.

Since its inception, the focus of the team has been on the following activities:

- Establishing a base-line "as-is" process model to support the development of a BPR Business Case Analysis (BCA).
- Providing test data to BPR pilot sites and an interim solution for the acquisition, integration, and delivery of data.
- Establishing standards for the format and structure of geospatial data.
- Testing and prototyping various AID methods, including burden sharing, data transformation, digital delivery, and data modeling.
- Planning long term AID strategy.
- Soliciting and documenting feedback from pilot site and other USDA business personnel.
- Conducting best practice forums to obtain lessons learned from similar efforts.

2.2 Purpose

The purpose of this document is to present the recommendations of the Data AID Project Team, document the activities which led to their determination, and establish a framework for the implementation of the recommendations. The recommendations identified in this strategy provide problem statements, case studies, and lessons learned from the project. Processes, standards, testing results, and costs have been used to formulate our recommendations. These recommendations provide specific implementation guidance to those implementing the recommended geospatial data AID plan.

2.3 Scope

It is the scope of this strategy to address the acquisition, integration, and delivery of geospatial information and to describe the activities of the Data AID Team and how each of the AID processes function. The activities within each of these AID processes lay the foundation for the recommendations and management implications presented later in this Strategy Plan.

2.3.1 Geospatial Data

Geospatial data is the subset of information that explicitly references a geographic position on the earth. Geospatial data typically contains a graphic component (point, line, polygon, pixel), and a tabular component, commonly referred to as attributes.

2.3.1.1 Acquisition

This document outlines the strategy for geospatial data acquisition, which is the process of gathering geospatial information meeting organizational business needs. These data assets can be collected internally (soil surveys) or externally (elevation models). As a result, the scope of the acquisition strategy must be comprehensive to include commercial procurement, partnerships, cost-sharing, and internal development.

2.3.1.2 Integration

Geospatial data integration is a critical process for providing geospatial data derived from a variety of sources and used to meet varied business needs in a seamless, plug-n-play fashion. The Data AID Team has developed geospatial data standards that address the format and structure of the geospatial information.

2.3.1.3 Delivery

The scope of the SCI covers a widely distributed network of field, state and national locations requiring access to geospatial data. Providing the right data, anywhere, anytime, to anyone in a timely manner while maintaining appropriate security measures, poses a major challenge. This strategy provides recommendations to meet these goals.

2.4 Acknowledgements

This Geospatial Data AID National Implementation Strategy Plan was produced by the Geospatial Data AID Team as the final report of its work from May 1998 to August 1999. The active members of the Geospatial Data AID Team are:

- Bruce Finch, FSA/APFO, Salt Lake City, (Co-Project Manager)
- Steve Nechero, NRCS/NCGC, Ft. Worth, (Co-Project Manager)
- Wendall Oaks, NRCS/ITC, Ft. Collins, (Co-Project Manager)
- Glenn Bethel, FSA-NHQ, Washington D.C.
- Elizabeth Cook, NRCS/Lincoln University, Missouri
- Rodney Johnson, FSA/APFO, Salt Lake City
- Hugh McGirt, FSA/APFO, Salt Lake City
- Jim Marron, NRCS/WCC, Oregon
- Fred Ringer, FSA-OR, Oregon
- George Rohaley, NRCS-NHQ, Washington D.C.
- Travis Rome, NRCS-KS, Kansas

- Lee Stieger, FSA, Kansas City
- Kent Williams, FSA/APFO, Salt Lake City

The Team is grateful to all of the Pilot site Service Center staff who provided us with valuable feedback regarding the delivered geospatial data. We would like to thank the staff of the Paola Service Center for spending the time to educate the team on the normal business needs of their Service Center.

We would also like to thank our Project Sponsors Diane Sharp, Ron Lauster and Maurice Mausbach, and our Reference Team members David Anderson, Randy English, Rob Vreeland, Greg Johnson, Emil Horvath, Phil Pasteris, David Tessier (Science Applications International Corporation), Tom McCarty (Science Applications International Corporation), Nicole Soltyka (Science Applications International Corporation), John Lee (Unisys), Jean Maloney (Unisys), Randy Frosh (Unisys) and John Evans (MIT). Also, thanks to all of the production staff at NCGC and APFO, especially Dwain Daniels, Elaine Ortiz and Ron Selph for all their help during the Pilot site support process.

Finally, we would like to thank Dennis Lytle (BPR Program Manager and Project Coordinator), W.R. Folsche (NRCS-NCGC) and Ron Nicholls (FSA-APFO) for their guidance throughout the project.

3 ACRONYMS

AID	Acquisition, Integration, and Delivery
APFO	Aerial Photography Field Office
BLM	Bureau of Land Management
BCA	Business Case Analysis
BPR	Business Process Re-engineering
CARAA	Conservation Area Resource Assessment Analysis
CCE	Common Computing Environment
CD	Conservation District
CD-ROM	Compact Disc Read Only Memory
CIR	Color Infrared
CFIP	Community Federal Information Partnership
CLU	Common Land Unit
COE	Corp of Engineers
COTS	Commercial Off-the-Shelf
CRADA	Cooperative Research and Development Agreement
CRP	Conservation Reserve Program
DBMS	Data Base Management System
DOQ	Digital Orthophotography Quadrangle
DOQQ	Digital Orthophotography Quarter Quadrangle
DOT	Department of Transportation
EDC	EROS Data Center
EPA	Environmental Protection Agency
EROS	Earth Resources Observation Systems
ESRI	Environmental Systems Research Institute, Inc.
FAC	Food and Agriculture Council
FAS	Foreign Agriculture Service
FEMA	Federal Emergency Management Agency
FGDC	Federal Geographic Data Committee
FOCS	Field Office Computing System
FSA	Farm Service Agency
FWS	Fish and Wildlife Service
FTE	Full-time Equivalent
FTP	File Transfer Protocol
FY	Fiscal Year
GAP	Gap Analysis Program of the USGS Biological Resource Division
GIS	Geographic Information System

HEL	Highly Erodible Land
I/O	Inter-Operability
IT	Information Technology
LAN	Local Area Network
MRB	Management Review Board
MIT	Massachusetts Institute of Technology
MDOQ	Mosaicked Digital Orthophotography Quadrangle
MrSID	Multi-resolution seamless image database; a raster data compression method
NAD83	North American Datum 1983
NASIS	National Soil Information System
NCGC	National Cartography & Geospatial Center
NCSS	National Cooperative Soil Survey
NDOP	National Digital Orthophoto Program
NIMA	National Imagery and Mapping Agency
NHD	National Hydrography Dataset (USGS)
NHQ	National Headquarters
NOAA	National Oceanic and Atmospheric Administration
NRCS	National Resources Conservation Service
NSDI	National Spatial Data Infrastructure
NSGIC	National States Geographic Information Council
NRI	National Resource Inventory
NWI	National Wetlands Inventory
PC	Personal Computer
PLANTS	Plant Information Database
POC	Point of Contact
RDBMS	Relational Data Base Management System
RPA	Risk and Productivity Assessment
RD	Rural Development
SC	Service Center (USDA)
SCI	Service Center Initiative
SCIT	Service Center Implementation Team
SSURGO	Soil Survey Geographic Database
SWCH	Snow, Water, Climate, & Hydrology System
TVA	Tennessee Valley Authority
TSSDS	Tri-Service Spatial Data Standards
URL	Uniform Resource Locator

USDA	United States Department of Agriculture
USFS	United States Forest Service
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WAN	Wide Area Network
WCC	Water and Climate Center

4 ACTIVITIES

4.1 Overview

This section describes the activities on which the Data AID Team focused in order to accomplish its charter. Briefly summarized, the activities included:

- Review guiding documents.
- Study existing As-Is situation.
- Conduct best practice reviews with federal, regional, and state government agencies.
- Perform pilot studies of all aspects of Data AID.
- Review successful and non-successful aspects of the pilot with Service Center personnel.
- Development of this document.

4.2 Study of As-Is Processes

Currently, efforts by non-BPR Service Center sites to support analog, and in limited cases digital geospatial data use are disparate, inefficient, and not scalable for enterprise implementation. Service Centers spend a good deal of time acquiring, reformatting, and translating geospatial information prior to being able to use it to service customers.

As an example, each Service Center agency collects, records, and evaluates geospatial data on separate paper bases. These source materials often vary in date, scale, projection, and quality causing confusion among customers and staff. When a change takes place on a farm the customer usually reports the change to one agency, which makes the update to their records. Since the change may not be communicated to all agencies, inconsistent data sets result.

The Data AID Team evaluated the existing agency infrastructure to support analog and digital geospatial data acquisition, integration, and delivery. There are a burgeoning number of Service Centers making some use of digital geospatial data, and for these, the AID tasks have largely fallen to state GIS staff. These efforts suffer from lack of national consistency and inadequate staff and tools to meet the demands. The review also considered legacy analog processes, and these should be replaced with digital solutions where appropriate. The evaluation included:

- Ordering hardcopy photography.
- Overlaying geospatial information for reference or analysis.
- Photo reducing or enlarging of information for varying scales.
- Photocopying of data for multiple business uses including reconstitution, planning, customer maps, etc.
- Transferring information from older to newer photography.

- Sharing geospatial information between USDA locations, federal, state, and local agencies and the general public.

4.2.1 Hardcopy Photography

Photographs or images that have become worn are reordered as needed. Because the photographs are usually from many different dates and times of year, when two adjacent photographs are displayed together, they often do not appear to align. Problems include:

- Water depth in lakes and ponds varies seasonally, causing the outline of the water body to misalign.
- Crops, crop maturity, and land management practices vary causing field boundaries to misalign.
- Roads may not align perfectly as they cross the photo border.

This presents a very large confidence problem for the Service Center customer, which should not be underestimated.

4.2.2 Common Land Units

Currently, all Service Center agencies maintain a wide array of information related to land units. This information is fragmented among paper documents and computer systems (i.e. System 36, FOCS (Field Office Computing System), etc.). Whereas each agency may delineate, manage and track land parcel information in a unique manner, there is a subset of common information that the partner agencies use. This subset of information is referred to as the Common Land Unit (CLU). Currently, land units are manually drawn on hardcopy photographs stored in each Service Center. A master copy of the photograph is stored along with copies of individual sections used for program or agency specific business. For example, FSA staff typically delineate farm, tract and field boundaries on 1:7,920 scale hardcopy photos and annotate the drawings with identification numbers, acres, highly erodible land (HEL) status and other relevant program related notes such as allotments, program payments, and crop reports.

NRCS staff typically uses the base maps to develop site specific conservation plans and engineering drawings. These drawings contain a variety of information related to the physical activity on the land. When FSA makes a change (reconstitutes the CLU), NRCS modifies their documents affected by the change, requiring several communications between agencies and duplicated efforts.

Finally, RD staff use the hardcopy photographs and the delineated land parcels as a basis for rural housing development. In summary, the current process is cumbersome, does not capitalize on data sharing opportunities, and results in duplication and error.

4.2.3 Soils

The National Cooperative Soil Survey (NCSS) is a collaborative effort between federal and non-federal partners to document the soil resources of privately held land in the U.S. NCSS has traditionally generated soil surveys for county-based areas using non-rectified, photo mosaics. However, digital orthophotography is now being used to ensure the spatial and locational integrity of the soil delineations and enhance the quality of image referencing. The soils layer is considered one of the four critical data layers for Service Center implementation.

The NCSS and NRCS have undertaken efforts to generate digital spatial and attribute soils data meeting Federal Geographic Data Committee Standards (FGDC). This data set is referred to as State Soil Survey Geographic Data (SSURGO) and is considered the highest quality, most detailed, nationally consistent, soil data set available for the U.S. As of August 1999, some 656 SSURGO data sets are available, of the roughly 2,500 areas identified as critical for national planning. Of these 656, roughly half have been compiled to DOQ, thereby facilitating full digital integration at the field level. The remaining 50% were compiled to analog orthophotography quadrangles and may present integration challenges before they can be fully used at the field level for planning. As a greater number of soil surveys are compiled to DOQ in preparation for digitizing, more surveys will be coincident with DOQ and therefore require minimal if any integration efforts.

4.2.4 Legacy Systems (non-spatially delineated)

A number of existing USDA data sets have spatial links, but lack geospatial delineations. For instance, the PLANTS database may contain references to the state and counties where a particular species can be found. However, there is no map showing the distribution of the plant. In addition, legacy systems vary in physical location, format, compression algorithm, and delivery mechanism. The Geospatial Data AID Team has studied how to best associate geospatial data with the existing tabular data and how to improve delivery of the data by providing a single access point to that data. There is no plan to replace these systems, simply enhance their usefulness to business applications and improve access.

Application Name	Functional Description of the Application
National Soils Information System (NASIS)	NASIS is used to collect and manage county and national soils survey information. A subset of this database is required by Service Center applications in support of the current legacy system FOCS as well the future Customer Service Toolkit (CST). Geospatial products at the county level are available to spatially enable this database.
Plant Information Database (PLANTS)	The PLANTS database is a repository of information on plant and crop species and their agronomic suitability for resource conservation use. In addition to the dependencies by other applications such as soils, plant information is present in Service Centers as a part of the legacy FOCS system. National maps to the county level are available.
Snow, Water, Climate, Hydrology (SWCH)	The National Water and Climate Center maintains a data repository for information related to Snow Survey and Water Forecasting activities. The FOCS Climate module provides basic data for use in conservation planning and other program delivery activities.

Table 4-1 Legacy systems

4.3 Best Practices and Outreach Efforts

The Team reviewed and evaluated “best practices” used by other governmental agencies and organizations that have experience with geospatial data acquisition, integration, and delivery on a large scale. The first session was held with the Corps of Engineers (COE) in Omaha, Nebraska on August 19, 1998. The second session was held with numerous federal agencies at the USDA Beltsville Office Complex in Beltsville, Maryland, on November 18-19, 1999. The discussion topics and lessons learned from both of these meetings are listed below:

4.3.1 Corps of Engineers Best Practice Session

On August 19, 1999 a Best Practice Session was held with the Corps of Engineers (COE) in Omaha, Nebraska. During this session the background of the COE Districts, their responsibilities, the history of GIS development in the Omaha COE, and technical issues in GIS implementation were discussed.

The COE GIS system is UNIX based with the Samba application software being used for network file system access between the PCs (DOS/Windows) and the UNIX system. The COE Local Area Networks (LANs) have a 100Mb-transmission rate. An Informix database is linked to the GIS system to store the tabular data. Most of the linked data is point (X, Y coordinates) data. COE does not use a Wide Area Network (WAN) to

network their sites. Each LAN is operated stand-alone with the data exchanged by tape, CD-ROM, and through the Internet.

The COE uses ArcView and maintains their files in an ESRI Shape file format. COE has various customized ArcView extensions and two primary types of applications:

- General purpose applications which allow users to search for data themes and perform automatic re-projection of geospatial data
- Specific purpose applications, such as Flood Risk Delineation Modeling

Below is a list of topics discussed with the COE:

- Training and End User Support
- Data Acquisition and Management
- File Format and Storage
- Software Customization
- Metadata

Training and End User Support

COE indicated training and timely help desk support are critical to a successful deployment and implementation. Relying on technical support from the system vendor was inadequate, due to slow responsiveness. System and network speed was critical to GIS system implementation and acceptance by the end user community.

Data Acquisition and Management

COE debated the pros and cons of acquiring data from an outside vendor, versus developing the data themselves. Purchasing data is less expensive but there will be overhead costs to integrate the data. COE-developed data integrated more seamlessly, but production costs were much higher.

COE is looking into future use of 1-meter DOQs, and they are using USDA soils data.

File Format and Storage

COE recommends all themes used for the same project be consistently formatted and standardized to eliminate distortion and alignment issues. They stressed the importance of training the end user in data limitations.

Software Customization

COE found that customized extensions/applications must be developed in order to meet the specialized needs of the end user community.

Metadata

Geospatial data sets generated at COE should adhere to Tri-Service Spatial Data Standards (TSSDS) for metadata. The body of TSSDS standards represents a single comprehensive data model for Air Force, Army, and Navy installations and the COE. Designed to compliment FGDC standards, TSSDS address issues surrounding

documentation of more detailed data sets. Representatives from COE indicated that the TSSDS metadata standard is not being met at COE due to the amount of time required to fully populate standard metadata elements.

4.3.2 Beltsville Office Best Practice Forum

On November 18-19, 1998 a Best Practice Forum was held at the USDA Beltsville Office to share "best practice" ideas associated with the implementation of GIS technology and to identify and discuss "lessons learned."

The forum was jointly sponsored by the SCI Data Management Team #5 (GIS Standards) and the Geospatial Data AID Team. The following agencies and organizations attended the forum:

- Bureau of Land Management (BLM)
- Environmental Protection Agency (EPA)
- Federal Emergency Management Agency (FEMA)
- Federal Geographic Data Committee (FGDC)
- National Imagery and Mapping Agency (NIMA)
- National Wetlands Inventory (NWI)
- Tennessee Valley Authority (TVA)
- US Census Bureau
- USDA Service Center Agencies (FSA, NRCS, RD)
- US Fish and Wildlife Service (FWS)
- US Forest Service (USFS)
- US Geological Survey (USGS)

Comments were discussed and documented for numerous topics including training, data management, and GIS Commercial Off-the-Shelf (COTS) evaluation. The meeting minutes for the Best Practices Forum is included in Appendix E. Comments pertaining to geospatial data AID are identified below:

- Data Acquisition
- Data Integration
- Data Delivery
- Metadata
- General Comments

Data Acquisition

- The US Census Bureau spent \$300 million to capture 20 million polygons.
- Some agencies such as NOAA have Cooperative Research and Development Agreements (CRADAs) with contractors to maintain value-added data.

- USGS indicated the source of some data is 10 years old and updating older data while mapping new areas for the first time creates a strain on their resources.
- TVA welcomed the opportunity to share and/or pay for the USDA DOQ data, as well as the data produced by the USFS, FEMA, and other agencies.
- USGS indicated APFO's DOQ quality control is valuable since USGS can not afford to perform more than a 10% random sample check.
- The USDA county prioritization for GIS implementation should be coordinated with partners and cooperators.
- Local governments are developing large-scale accurate data that could be used in place of the data being developed at the federal level. The federal government should not spend limited resources to duplicate these efforts.

Data Integration

- The USFS has collected data for the past 15 years on a regional basis and is now addressing integration issues.
- USGS cautioned against providing data in all formats and suggested this be the role of after market vendors.

Data Delivery

- The USFS Geometronics Service Center in Salt Lake City is targeted to become a national warehouse and data clearinghouse for USFS.

Metadata

- USGS has not had much success finding usable metadata collection or management tools.
- All of the participating agencies are capturing metadata but not necessarily completely or consistently. Metadata capture tools are not readily available.
- FWS users would like to have metadata to catalog paper maps and to track digital data. So far however, this is a work in progress.
- USGS advised USDA to include a metadata tool in the Enterprise GIS requirements to ensure contractual leverage during procurement.

General Comments

- BLM commented that implementing a large project detracts from the agency's real mission. Consequently, BLM will not be implementing any more large projects. BLM added that project leaders should be the drivers of these types of projects, not the Information Technology (IT) group.
- TVA indicated that communication and training were the most difficult aspects of implementing GIS technology.
- The US Census Bureau advised the group to remove all options except success to increase commitment, and cautioned against overanalyzing the intricacies of implementation. They stressed the importance of deploying the tools.
- EPA tried deploying desktop GIS, but the training was not successful and a Web-based system proved to be a cheaper alternative.

4.4 Pilot Site Support Activities

To support current BPR initiatives, test alternatives, and increase our knowledge of performing geospatial data AID, nine pilot Service Centers covering eleven counties were initiated. This section details the experiences, lessons learned, and Data AID policies and procedures used to provide geospatial data AID support for the pilot sites.

The SCI pilot support process primarily involved the following activities:

- Providing "value added" geospatial data to the SCI pilot sites.
- Providing hotline assistance to the pilot sites regarding the data provided.
- Assisting the SCI BPR projects with the definition of their geospatial data requirements and acquiring, integrating, and delivering the data to the applicable pilot site(s).
- Defining the geospatial directory and file naming conventions to be used at each pilot site.

The organizations supporting the pilot sites are the National Cartography & Geospatial Center (NCGC) in Ft. Worth, Texas, the Aerial Photography Field Office (APFO) in Salt Lake City, Utah, and the Inter Operability Lab (I/O Lab) in Beltsville, Maryland. NCGC acquired and integrated the soils and demographic data, as well as the various non-critical data themes. APFO acquired and integrated the Digital Orthophoto Quadrangle (DOQ) and Common Land Unit (CLU) data. The I/O Lab coordinated the receipt of all pilot site geospatial data from NCGC and APFO and then delivered the data to the pilot sites. NCGC worked with state GIS staff in the pilot states to supplement national themes with state and locally derived themes when appropriate.

The acquisition, integration, and delivery of geospatial data to the nine pilot sites began in July 1998. Valuable lessons were learned while working with the first few pilot sites; consequently numerous adjustments to the content of the data and its structure were made. Having an identical structure in each site was extremely important for proper support of the sites. Therefore, in January 1999 a data refresh process was started providing all pilot sites with geospatial data in a consistent directory and file structure. This process was completed in June 1999. See Figure 1 for the locations of the pilot sites.

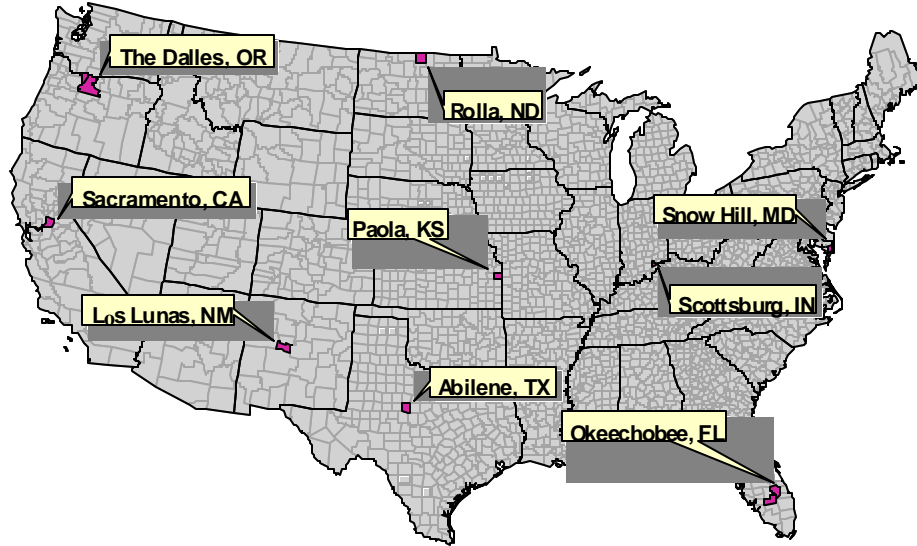


Figure 1 – USDA Service Center Initiative Pilot Service Center Locations

4.4.1 Data Themes

The *USDA Service Center Geographic Information System (GIS) Strategy* document, April 17, 1998, lists 19 geospatial data themes and sample business needs for those themes. The document does not detail the scale, database fields, source, or other requirements for those themes. However, the document does state that we must deliver the “best available data.” Therefore, at the start of the pilot, the Team acquired, integrated, and delivered the best available data as identified in the *GIS Strategy* document. Although the Team learned a great deal about available federal and state geospatial data, it is necessary to define data needs according to application requirements and not according to data availability. Since some applications are developed and maintained locally, data needs must be defined according to application specifications. In order to ensure that data located at Service Centers support business needs, Service Center staff and other BPR project coordinators should work together to identify specific data requirements. The National GIS Coordination Team should coordinate with both the Service Centers and the BPR teams to ensure that Service Centers receive the data that truly meet their requirements and those of the BPR initiatives. BPR teams have started to analyze their business requirements and have requested specific geospatial data sets. However, the issue of defining data requirements according to applications still exists and must be addressed prior to national implementation. Clearly defined data requirements will ensure appropriate geospatial products are selected to meet the business needs of the Service Centers.

4.4.2 Critical Data Themes

The four critical data themes defined in the *USDA Service Center Geographic Information Strategy* document are Digital Orthophoto Quadrangles (DOQ), Common Land Units (CLU), soils (SSURGO), and demographic data. Consequently, the primary focus of the Data AID Team was to deliver the four critical data themes to the nine pilot sites. It is important that USDA maintain consistency among the geospatial data sets and their corresponding attributes across all Service Centers.

DOQs are produced through the National Digital Orthophoto Program (NDOP). NDOP is a national program that solicits cost sharing from USDA, other federal agencies, and state governments. NDOP attempts to meet the requirements of partners by defining a common set of specifications. USGS administers NDOP and distributes DOQs to their cost share partners.

The integration process involves mosaicking the DOQs to present a single, “seamless” image of a county. Visible seam lines between separate DOQs due to different dates of source photography within a county, misalignments between DOQs, and other factors, are minimized during the mosaicking process. Two products are generated from the mosaicked DOQs (MDOQs), a compressed mosaic of individual counties and a set of full-resolution imagery tiled according to 7.5-minute quadrangle boundaries. The compressed county mosaic provide county-wide views and facilitates applications involving laptop computers, while the MDOQs are used for digitizing CLU data, performed at either a FSA digitizing center, or by a data conversion contractor.

Soils data acquired for the Service Centers consist of spatial and tabular data. Tabular data are stored in a Microsoft Access database derived from the National Soil Information System (NASIS). This database consists of 49 data tables, 10 metadata tables, 12 frequently used queries, and 2 static metadata queries.

The U.S. Census Bureau develops demographic data. Demographics consist of spatial boundaries, such as Census Tracts, Minor Civil Divisions, and Voting Districts, and tabular data or attributes, such as population, income, and housing statistics.

Demographic data are reformatted as required by the Service Center GIS software and projected into a standard projection system. A number of derivative themes are generated, making GIS analysis easier for Service Center staff.

Section 5 details the source, scale and cost to acquire and integrate each of the four critical layers.

4.4.3 Non-Critical Data Themes

The non-critical data themes delivered varied by pilot site due to data availability. The best available data were not always nationally consistent. For example, detailed GAP (Gap Analysis Program of the USGS Biological Resource Division) land cover was not

delivered to all pilot sites because it is not available for some locations. In areas lacking GAP land cover, a more generalized and older nationwide data set was delivered.

All non-critical data themes were delivered based on v5.0 of the Directory Structure and File Naming Convention, developed by the Data AID Team, and projected into UTM NAD83 (North American Datum 1983) coordinates. This directory structure is included in Appendix D.

A list of the non-critical data themes delivered to each of the nine pilot sites is included in Appendix C. Section 5 details the source, scale, and labor costs to acquire and integrate each of these layers.

4.4.4 Determining Future Data Themes

The Data AID Team anticipates that business needs will be identified for non-critical data which are not presently available at the pilot sites and which are not listed in the original GIS Strategy document. As BPR project teams and Service Centers are educated in the use of GIS, they will identify new data themes and analysis needs. The Data AID Team will work with the BPR project team and others to identify and resolve data issues before additional data sets are developed. These issues are as follows:

- Required geospatial data scale.
- Required attributes (associated database information).
- Requirements of these data by other BPR teams to avoid re-development later.
- Budget for the data.
- Suitable source for the required data.
- Publisher name, location and Internet address.
- Source data steward and a USDA data steward responsible for the data process within the USDA.
- Delivery media options.
- Availability schedule.
- Possibility for partnering/cost share for data development, particularly at the state and local level.

4.4.5 Geospatial Data Directory Structure and File Naming Conventions

The geospatial data directory structure and file naming conventions evolved over the first six months of the pilot support process. This resulted in each of the initial pilot sites having a directory structure and file names that were unique per site (v1.0, v2.0, v3.0, etc.). In order to correct this situation the Data AID Team developed v5.0 as the standard to be implemented at all nine pilot sites. Consequently, as the remaining pilot sites were provided with data, and as the previous pilot sites were provided with re-delivered data, their directory structure and file names were updated to adhere to v5.0.

Consistent directory and file structure is very important in the support of the Service Centers. Therefore, we recommend standards for both and a rigorous change control procedure be established.

4.4.6 Pilot Data Refresh

As was the case with the directory structure and file naming conventions, there were inconsistencies in the type of data sent to each pilot site, as well as the format and projection used. Consequently the Data AID Team prepared a pilot site data refresh plan. The mechanism used to refresh the pilot data depended on the amount of data to be replaced, the availability of more current data, and the directory structure currently residing on the pilot server. The process for data refresh was as follows:

- Where the current directory structure was not consistent with v5.0, a back-up tape of pilot site data was prepared in the I/O Lab and shipped to the pilot site for loading. The back-up tape contained the v5.0 directory structure and all of the available geospatial data modified at the I/O Lab to adhere to the v5.0 file names. The Service Center staff replaced existing geospatial data with the v5.0 geospatial data from the back-up tape.
- When smaller size data sets containing more recent data and the current directory structure at the pilot site was consistent with v5.0, data was refreshed remotely from the I/O Lab. Only more recent data sets were refreshed.
- When larger size data sets containing more recent data and the current directory structure at the pilot site was consistent with v5.0, data was sent to the pilot site on a back-up tape. Only the more recent data sets were included on the tape.

These procedures highlighted the need for a higher capacity media and bandwidth to decrease the access time and cost of data refresh.

4.5 Review of the Pilot Data with Service Center Personnel

In addition to weekly review meetings with Service Center pilot staff, a survey was developed and distributed to all pilot sites. Each agency located at each site was asked to respond to the questions about geospatial data usage within their agency. All nine pilot sites responded to the survey. A total of 24 responses were received. Of the 24 responses received, two were from Conservation District (CD) Partners. Removing the CD responses from the return results yielded a pilot site return rate of 81.4 percent. The questionnaire highlighted a number of successes and failures of the pilot program. Key observations are as follows:

- The data themes identified as critical are only some of the data layers used most often at most sites. There is a need to re-evaluate what is critical based on these survey results. Feedback indicates that the hydrography, NWI, and DRG data sets are equally as popular or more popular than the “critical” demographic data set.
- Incomplete data sets are used less frequently.

- Data must be delivered for all counties served by each agency. RD's servicing territory often covers larger areas than FSA or NRCS, in which case the data was incomplete for RD.
- All agencies use the DOQs. Each agency uses additional layers with varying frequency.
- Sites have adapted to use GIS at different rates.
- Non-critical themes most frequently used are hydrography, NWI, Digital Raster Graph (DRG), and transportation.
- A significant portion of the pilot site agencies are not using the GIS data yet.

At each pilot site, at least one of the three agencies (NRCS, FSA, and RD) is using some geospatial data. The breakdown of geospatial data use by agency is detailed in Table 4-2. This information was obtained from the survey results when available and supplemented with information obtained from the Pilot Site Coordinator where information was not provided.

Agency	# of Sites Using	# Sites Not Using
NRCS	9	0
FSA	6	3
RD	2	7

Table 4-2 Pilot data use survey participants

Regardless of site or agency, the top three responses to each question are listed below in Table 4-3. Response percentages are included in parenthesis next to the corresponding data set that they represent. These percentages are based on all 24 responses received as of September 7, 1999.

Question	#1 Response	#2 Response	#3 Response
Critical Data Theme Used Daily, Weekly, or Monthly	DOQ (70.1%)	CLU (66.7%)	Soils(58.3%), Demographics (20.8%)
Non-Critical Data Theme Used Daily, Weekly, or Monthly	Hydrography (29.2%)	Wetlands, DRG (25.0%)	Hydrologic Unit, Transportation (20.8%)
Data not delivered but would be used	Flood Zones, Tax Parcels (16.7%)	Underground Utilities, Completed DOQ, CLU, and Contours (12.5%)	Crop Use/History, Land Use/Land Cover, Wetlands, T&E Species, Better Transportation and Hydrography (8.3%); Adjacent County Data, Nitrate Leach Index, Soil-Pesticide Interaction Index, Hydrologic Subunits, Wind, Evapo-Transpiration, CIR DOQ, Hazards, Rangeland, Stream Gauging Stations, NRCS Snotel and Water Forecasting Stations, Wells, Geologic Maps, Aquifer Maps, Irrigation Districts (4.2%)
Data delivered but never used nor for which a use can be foreseen	Unsure (62.5%)	Climate (12.5%)	Demographics, Government Units, Hydrologic Units, Transportation, Plants, Land Use/Land Cover (8.3%); CIR DOQ, Elevation (4.2%)
Do you use DEM or contour lines for elevation data?	Haven't used data, did not know difference, or did not answer question (91.7%)	DEM (4.2%)	Contours (4.2%)
Do you use MrSID compressed or TIFF DOQs?	Neither (33.3%)	MrSID (33.3%)	TIFF (33.3%)
Would you prefer DOQs by township or 7.5' quad?	Do not care or do not know (62.5%)	Township (29.2%)	Quadrangle (8.3%)

Table 4-3 Pilot data use survey results

4.6 Lessons Learned from the Pilot Support Process

Lessons learned from the pilot testing fall into four categories. The four categories are listed below followed by summaries of each.

- Data
- Process
- Customers
- Team

Data Lessons Learned

- Complete coverage of geospatial data at the scale needed for field level analysis does not exist.
- Even though some data layers meet national map accuracy standards, they do not align satisfactorily with the DOQ and DRG images.
- More communication is needed with data producers to acquire metadata and certified data to support GIS applications.
- Some Digital Orthophoto Quarter-quadrangles (DOQQs) delivered by USGS contain errors in positional accuracy, metadata information, and radiometry. Inspection and correction of imagery at APFO is critical for GIS applications in USDA Service Centers.
- The source materials for the CLU vary in quality, consistency, and completeness. There is a need for standardization of data collection in compliance with the CLU Handbook (FSA 8CM).
- Imagery for the production of DOQQs is continually becoming available. Therefore, it is necessary that the flying date of source materials match the flying date of ortho-imagery, so that the most current imagery can be provided to the Service Centers.
- There is a demand for Color Infrared (CIR) ortho-imagery in some Service Centers.
- Full resolution MDOQs are best for digitizing CLU.
- The ortho-image base must have the flexibility to be updated over time for specific areas within a county, resulting in a DOQ refresh for the affected counties.
- Excellent state and local geospatial data sources exist; however in some cases they are not documented and require significant resources to find, acquire, and integrate.
- State GIS staff know the availability, quality and geographic extent of state and local data to supplement nation-wide themes and their involvement is critical to providing the “best available” data to Service Centers.
- Very little FGDC-compliant or other metadata exists for geospatial data.
- The scale needed for field level analysis, such as 1:24,000, must be specified by the requesting agency.
- Because the service area often differs for each agency in a Service Center, data for the largest area must be delivered for it to be useful to all agencies. For example, RD offices often service larger geographic areas than NRCS or FSA.

Process Lessons Learned

- Data quality must occur early in database development to reduce problems with data revisions and maintenance.
- A timely and efficient process is needed to obtain revisions from data stewards after errors in a data set are found and reported.
- Contacts with knowledge and access to locally generated geospatial data need to be developed; in particular, the Service Center Implementation program needs to involve state GIS staff who are often knowledgeable of state and local data and partnering opportunities.
- Essential cartographic elements need to be created and approved for consistent national implementation.
- County mosaics are created more efficiently if groups of counties are processed in blocks. The production of the ortho-imagery mosaics, the digitizing of the CLU and the training of the CLU digitizing center staff must be coordinated for production and delivery of the CLU data.
- The Service Centers must be able to edit and update the CLU data with tools that ensure the topologic integrity of the data.
- DOQQ mosaic seam-line definition is inherently subjective, resulting in more than one interpretation of the appropriate seam-line placement.
- Hardcopy photography in use at the Service Center may be older than the photography used to create the DOQ. When a third party creates the CLU layer, it is required that the hardcopy source materials match the softcopy DOQ. Additionally, hardcopy may need to be provided to Service Centers with DOQ source photography that is older than 1994.
- The identification of a USDA data steward or non-USDA data steward, depending on applicability, must occur early in the process of data acquisition and integration. This is necessary to answer questions about the data, including aid in generation of metadata if it does not exist.
- A repository of the "official version" of the data should be established in order to refresh the data and avoid not knowing which version is "correct".
- A concrete process to add or alter data theme definitions should be established in order to prevent "theme creep" and "attribute creep" that may expand the size and scope of the geospatial data and their attributes outside of their original intent.
- Service Centers will inevitably wish to update data layers. There must be a process to avoid over-writing their modifications during a data refresh.
- Data generated at the Service Center will often need to be aggregated up to state and national data themes to meet other business needs of the agencies.
- Pilot Sites require hotline assistance to answer technical data questions.

Customer Lessons Learned

- GIS experience varies tremendously between offices.
- Data requirements other than the critical data themes vary between agencies more than they do between sites, although the sample size of nine is small.

- Clients require education on the personnel and resource costs of Data AID and storage.
- Clients require education on how to define their own data requirements.
- Client requirements can change quickly as technology changes and as they are educated.
- Unless data delivery is complete for a theme for all agencies at a Service Center, the data is not always useful to the client. It is easier for them to continue with their manual processes until the data are complete. This will hinder GIS use in the Service Centers.

Team Lessons Learned

- Smaller sub-groups can make faster decisions.
- The team leader should have this responsibility as a primary task, not as a "side" task.
- A clear definition of project objectives, boundaries, budget, and authority is required for the team to operate effectively.
- Meetings should be two days or less and focus on concise subjects.
- Teleconferences should occur only as often as necessary.
- It is important to maintain contact with the end-users of the data, either through BPR teams or directly, in order to understand the GIS literacy of the end-user and intended uses for the data.

5 GEOSPATIAL DATA STATUS

5.1 Critical Themes Status Maps

The maps in this section display the availability of the critical data themes with the exception of demographic. Demographic data available from the U. S. Census Bureau and the USDA (Agricultural Census) is currently available for all counties.

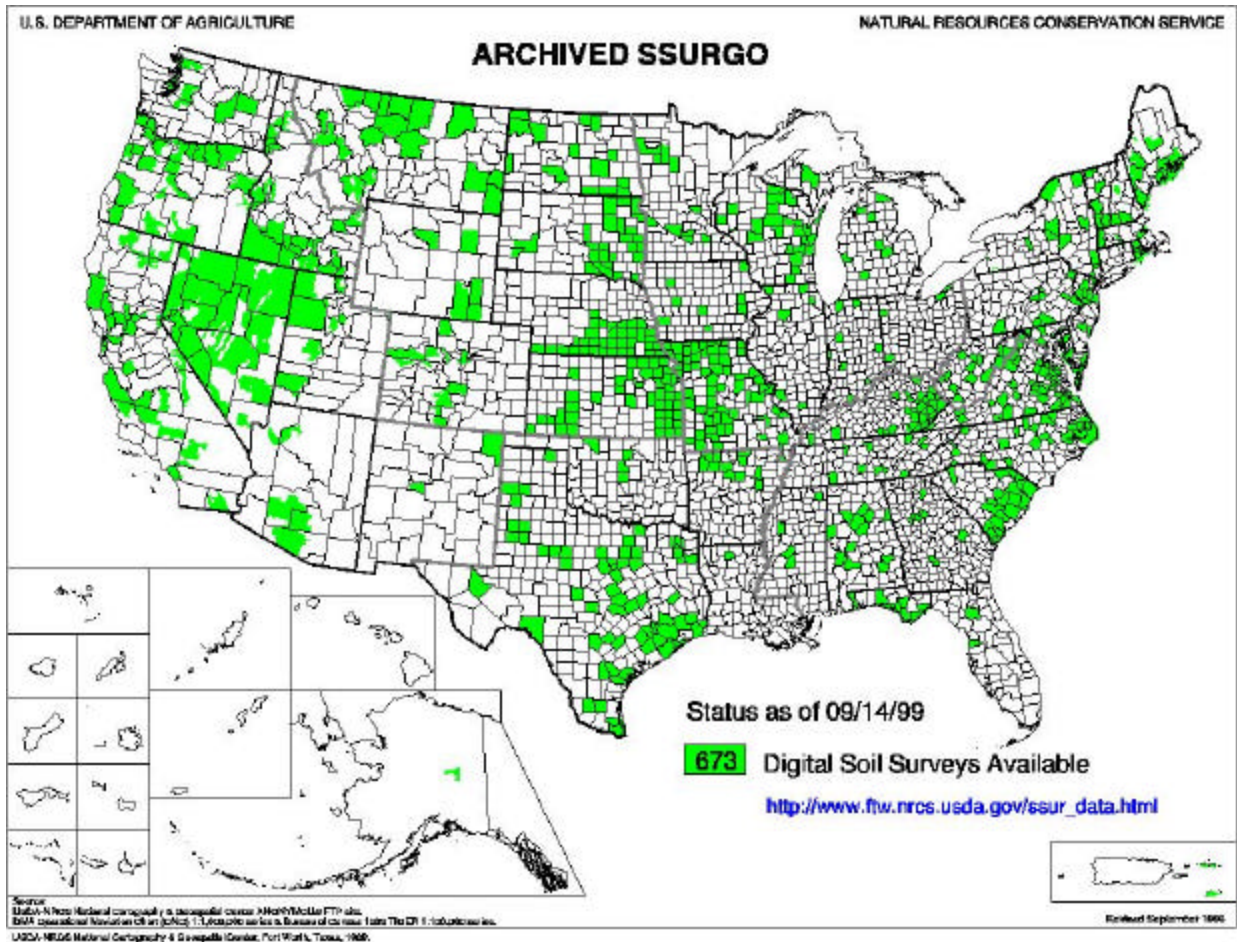


Figure 2 displays the areas of the country that have Digital Soil Surveys (SSURGO) available as of September 14, 1999. This status information is available in greater detail on the Internet at the URL, http://www.ftw.nrcs.usda.gov/ssur_data.html.

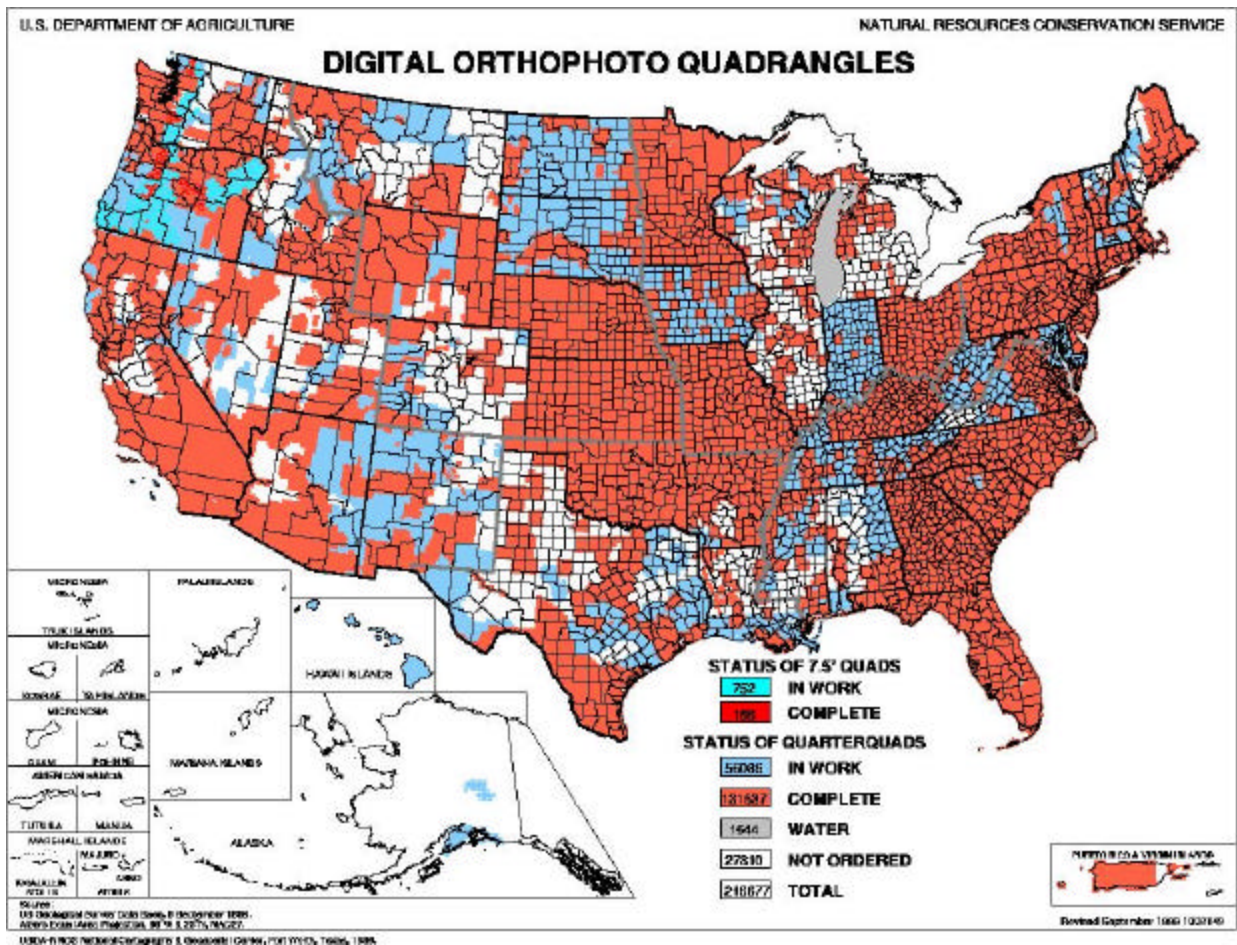


Figure 3 displays the status of USGS Digital Orthophoto Quadrangles throughout the country as of September 9, 1999. Several states have complete coverage. There are approximately 1,246 counties with 100-percent complete DOQ coverage.

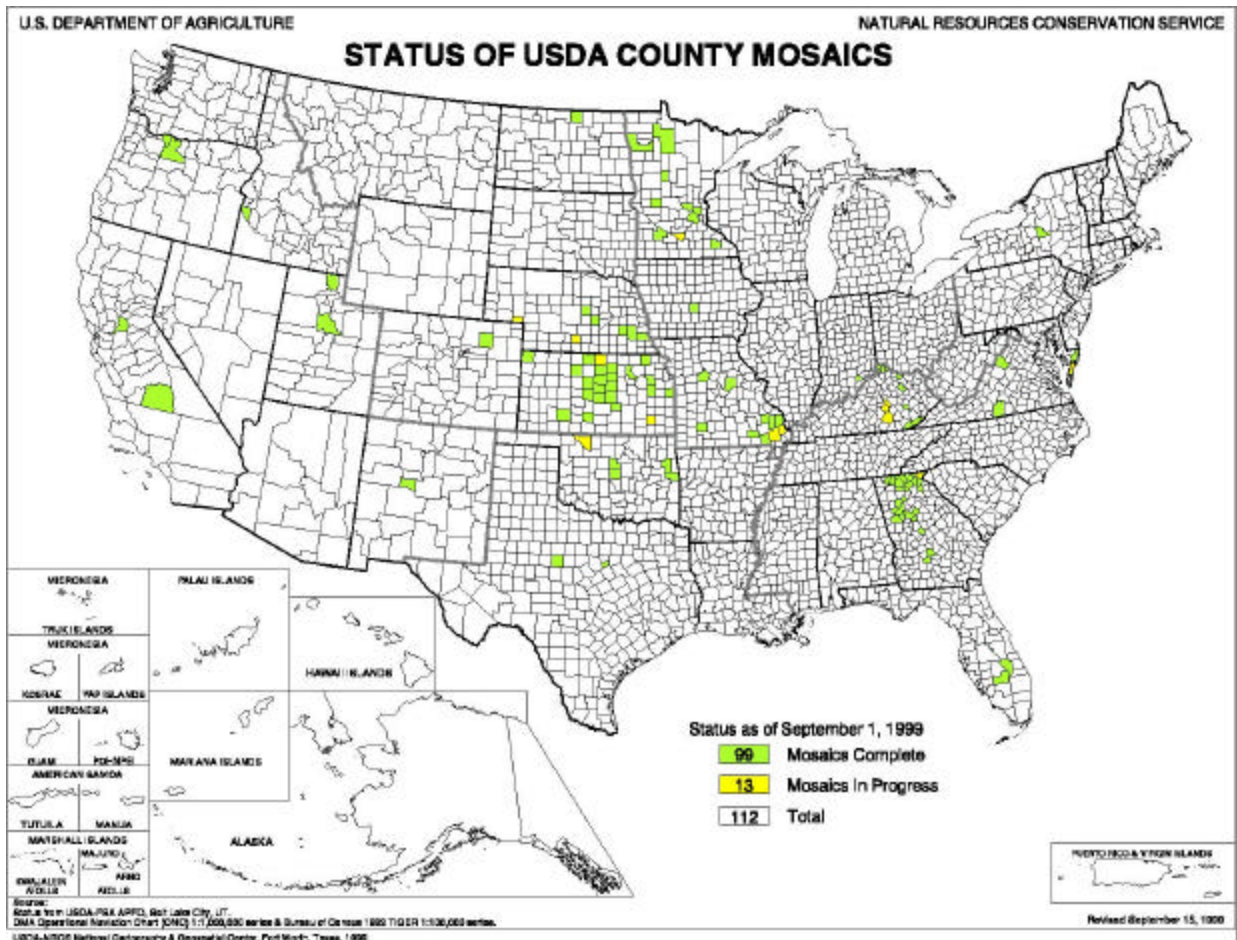


Figure 4 displays the counties that have had a mosaic of the Digital Orthophoto Quadrangles (MDOQ) prepared, or in progress as of Sept. 1, 1999.

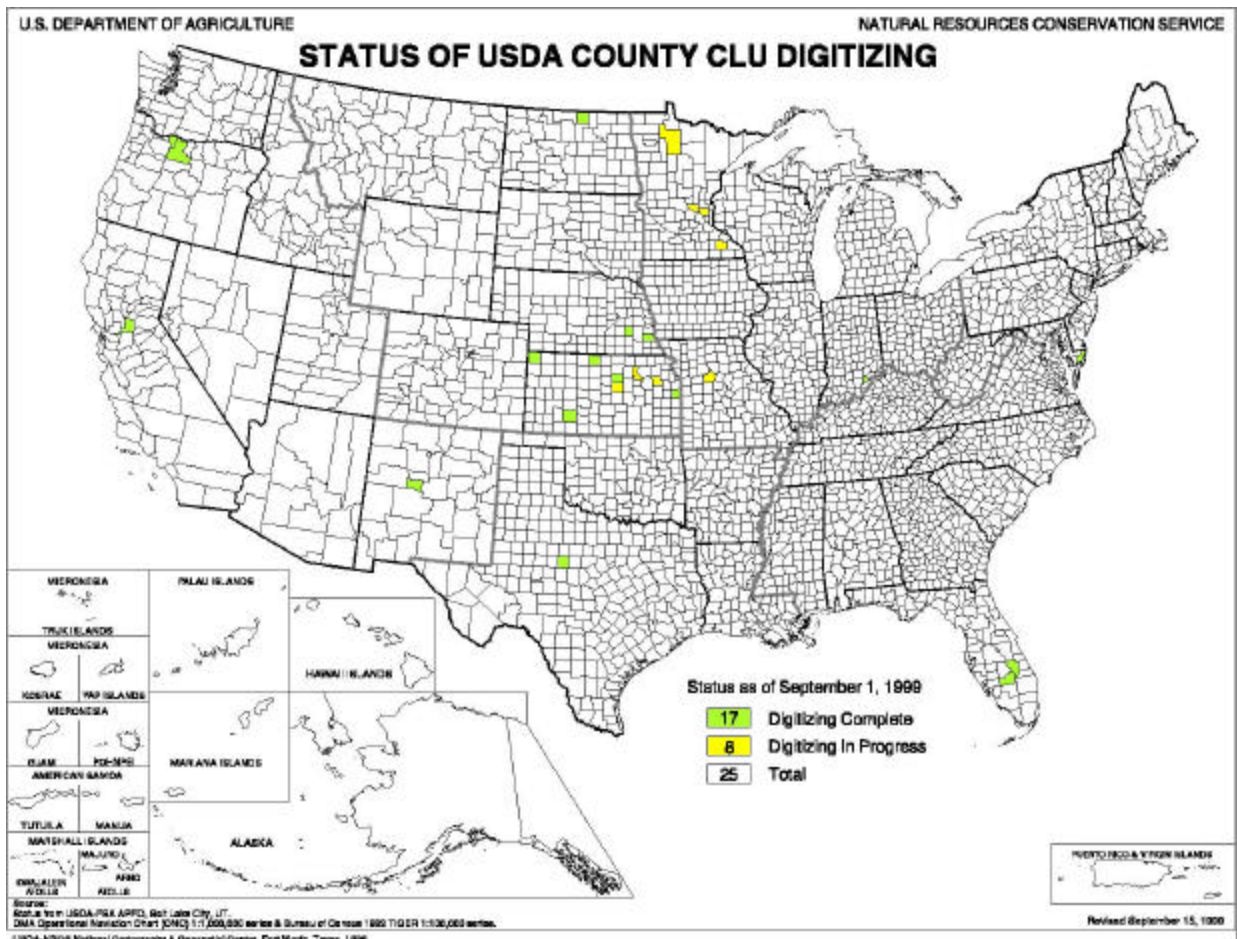


Figure 5 displays areas in the country that have had the Common Land Unit (CLU) data set prepared as of September 1, 1999. The CLU digitizing process is currently accomplished by twelve digitizing centers. APFO is performing a quality assurance inspection on the line-work for each digitized county. If line-work is not satisfactory, the CLU boundary is tagged to show up on a check plot as an area for the Service Center or digitizing center to examine for errors. Currently, the files are transferred to APFO from the digitizing centers via FTP. The files are being inspected for quality and subsequently metadata is created. Finally, a check plot is generated depicting the CLUs that were randomly selected for verification and which were actually inspected. The files (CLU, metadata, and checkplot) are subsequently written to CD-ROM using the AID v5.0 directory structure and distributed to the state FSA GIS Coordinators.

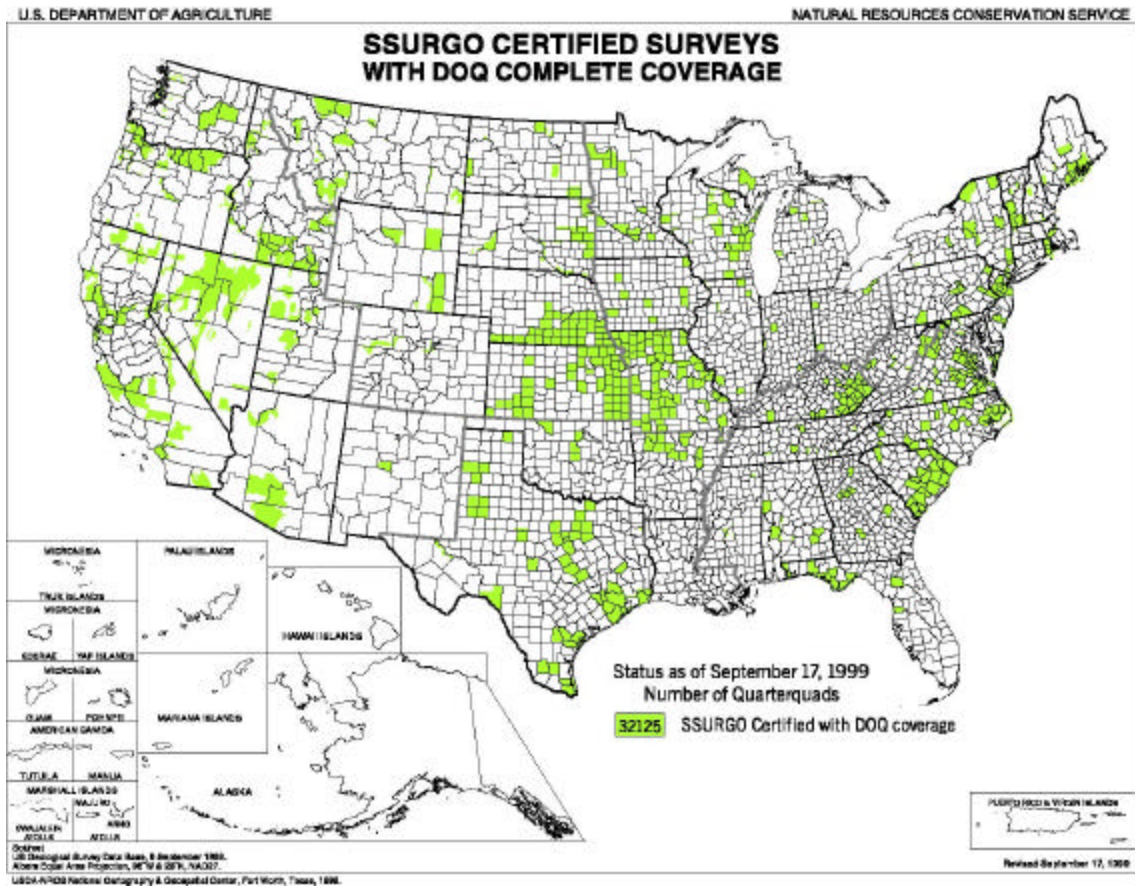


Figure 6 depicts areas where the availability of SSURGO certified survey data and DOQ coverage is coincident. This figure highlights the importance of reconciling the production schedules where the critical theme information is available.

5.2 Non-Critical Themes Status Maps

The following maps depict the production status of some of the non-critical data sets. Several of these data sets are obtained from FEMA and USGS.

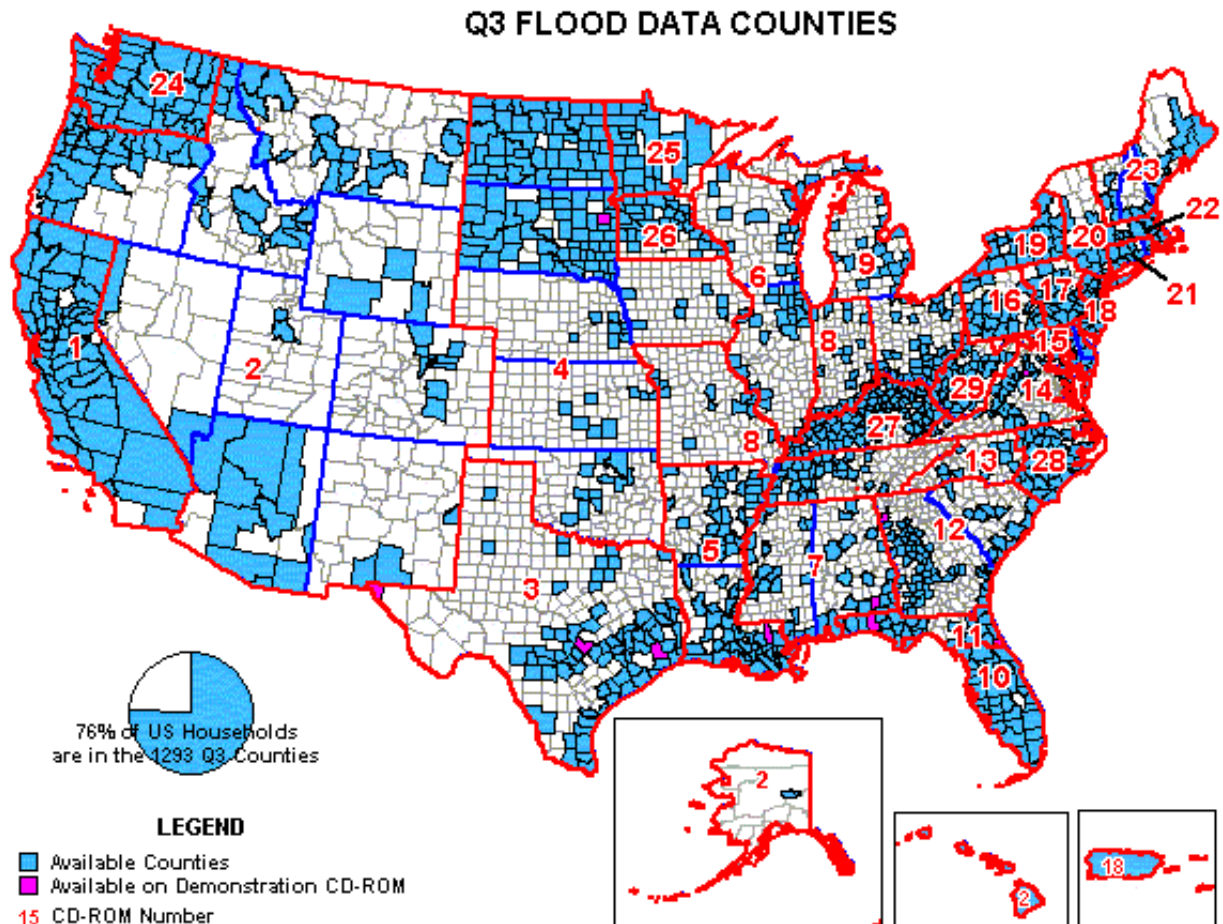


Figure 7 displays the Federal Emergency Management Agency (FEMA) national coverage of digital maps of flood hazard areas. This status map can be viewed on the Internet at this URL, <http://www.fema.gov/MSR/statemap.htm>. Digital Q3 Flood Data are now available for more than 1,250 counties.

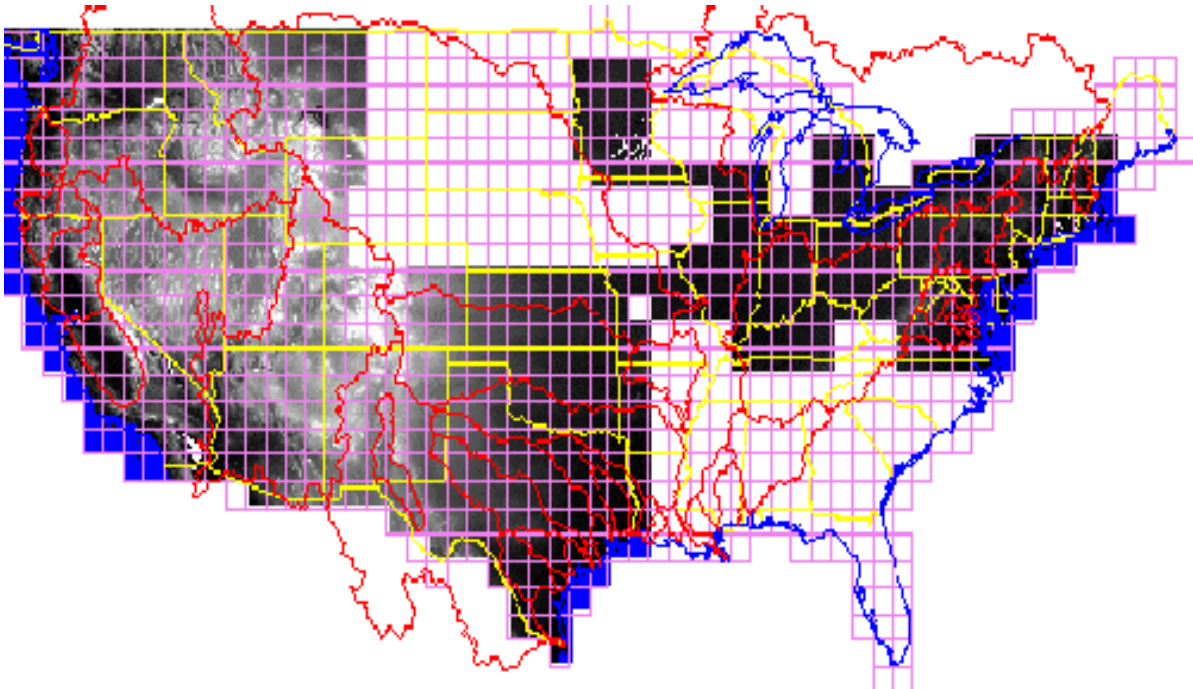


Figure 8 displays the extent of the 7.5-minute Digital Elevation Model (DEMs) conversion to the National Elevation Dataset (NED) mosaic for the conterminous 48 states. This status map can be viewed on the Internet at this URL, <http://edcwww2.cr.usgs.gov/umap/ned/ned.html>.

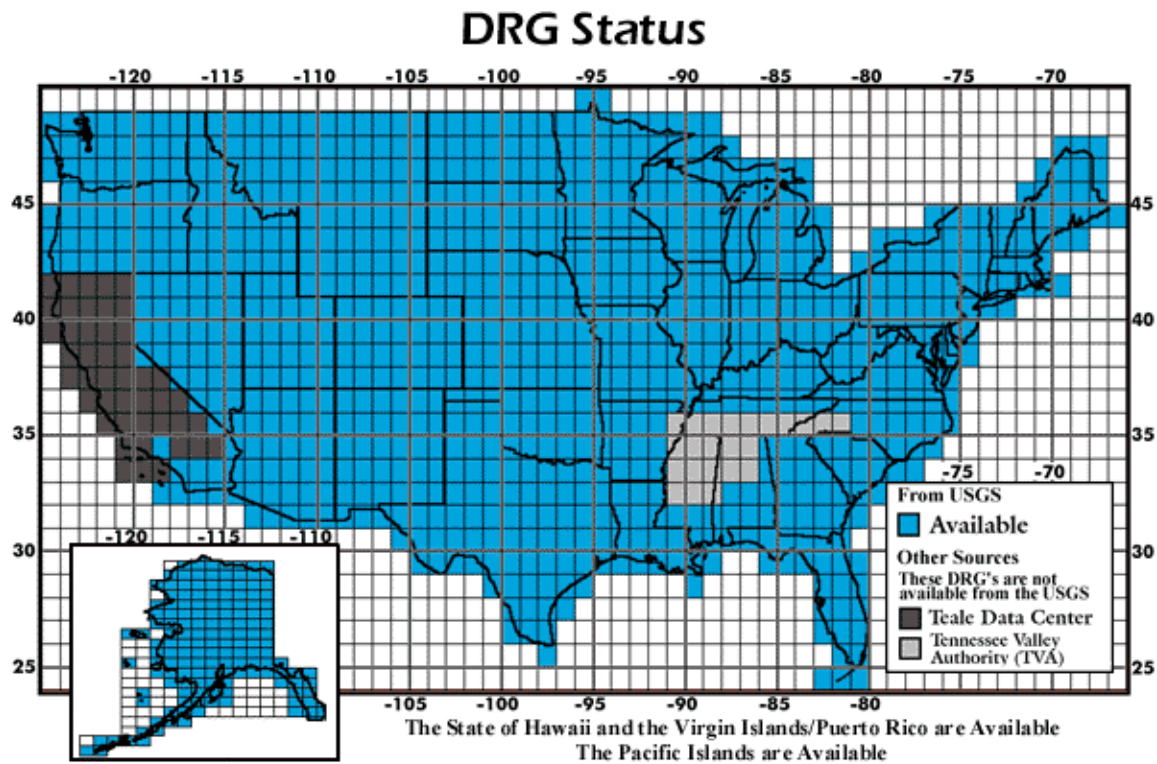


Figure 9 displays the areas where the USGS Digital Raster Graph (DRG) are available. This status map can be viewed on the Internet at this URL, http://mcmweb.cr.usgs.gov/status/drg_stat.html.

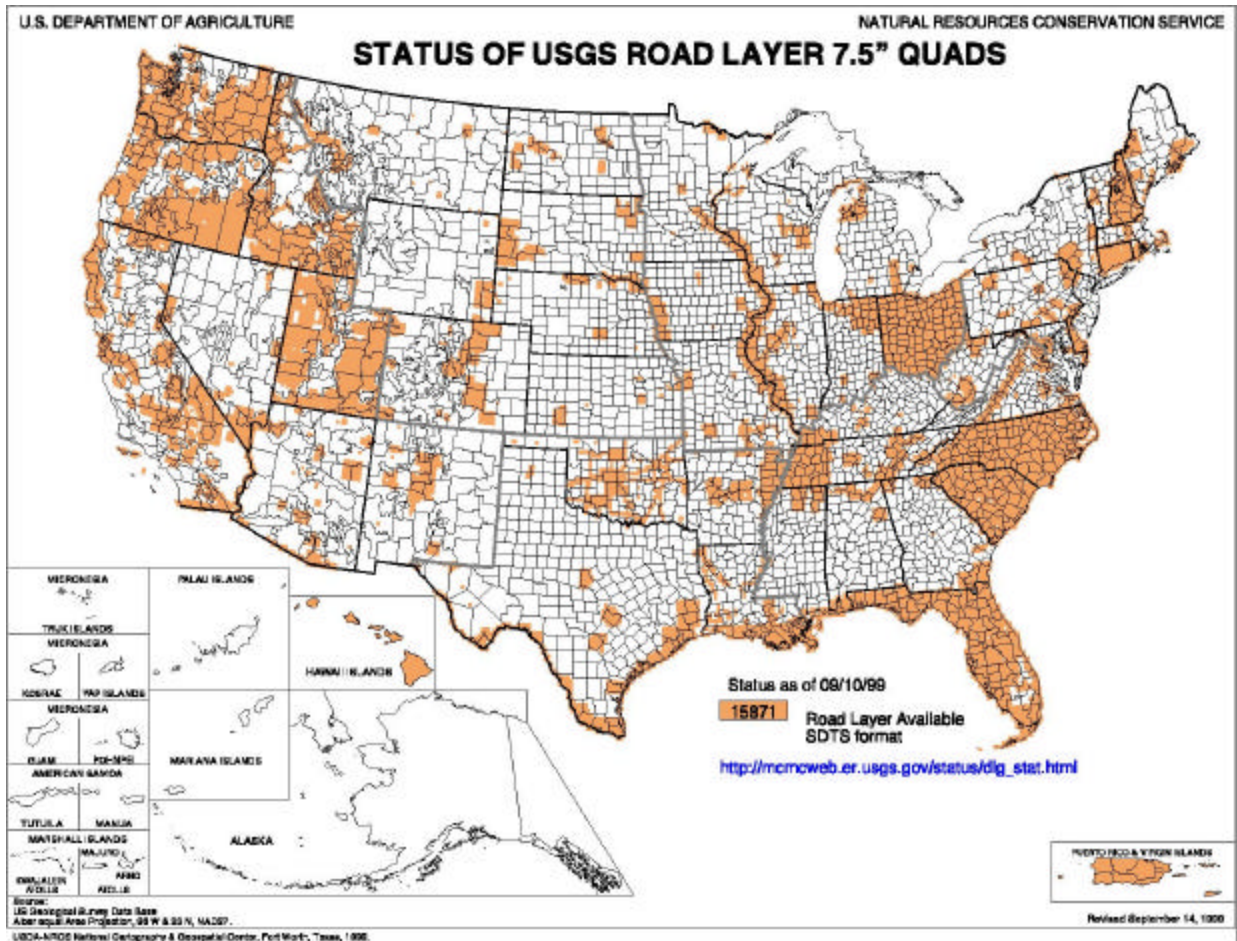


Figure 10 shows the status of the USGS road transportation data layer that is currently available for acquisition, integration, and delivery to Service Centers. This status information is available in greater detail for display at the state level over the Internet at the URL, http://mcmweb.er.usgs.gov/status/dlg_stat.html. A few states, Rhode Island, Connecticut, New Hampshire, Washington, Delaware, Oregon, Utah, Ohio, Florida, North and South Carolina have substantial or complete coverage at this time. This data layer can be produced when the 7.5-minute DRGs are prepared. A railroad transportation layer is also available.

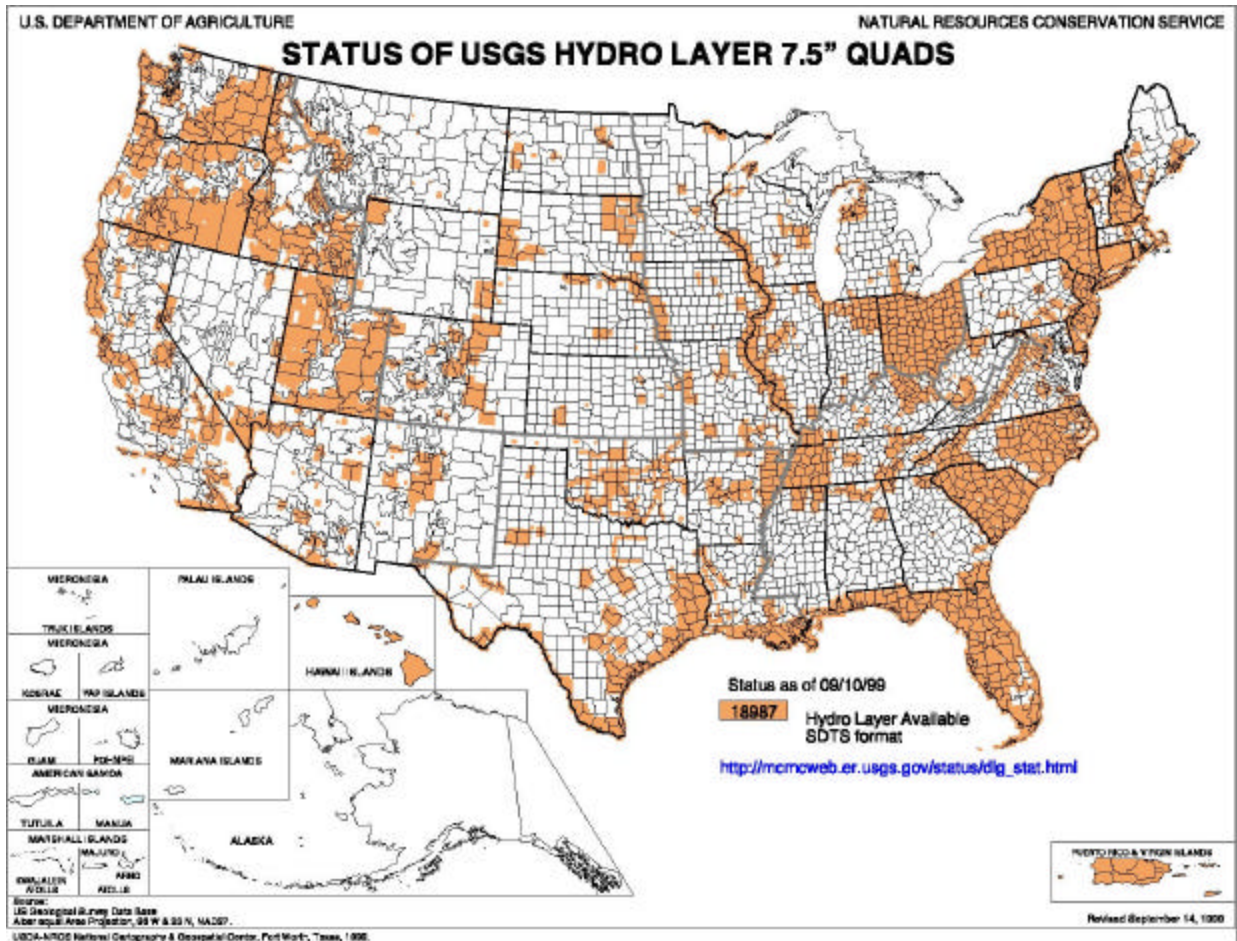


Figure 11 shows the status of the USGS hydrology data layer that is currently available for acquisition, integration, and delivery to Service Centers. This status information is available in greater detail for display at the state level over the Internet at the URL, http://mcmweb.er.usgs.gov/status/dlg_stat.html. A few states New York, Rhode Island, Connecticut, New Hampshire, New Jersey, Delaware, Washington, Oregon, Utah, Ohio, Florida, North and South Carolina have substantial or complete coverage at this time. This data layer can be produced when the 7.5-minute DRGs are prepared.

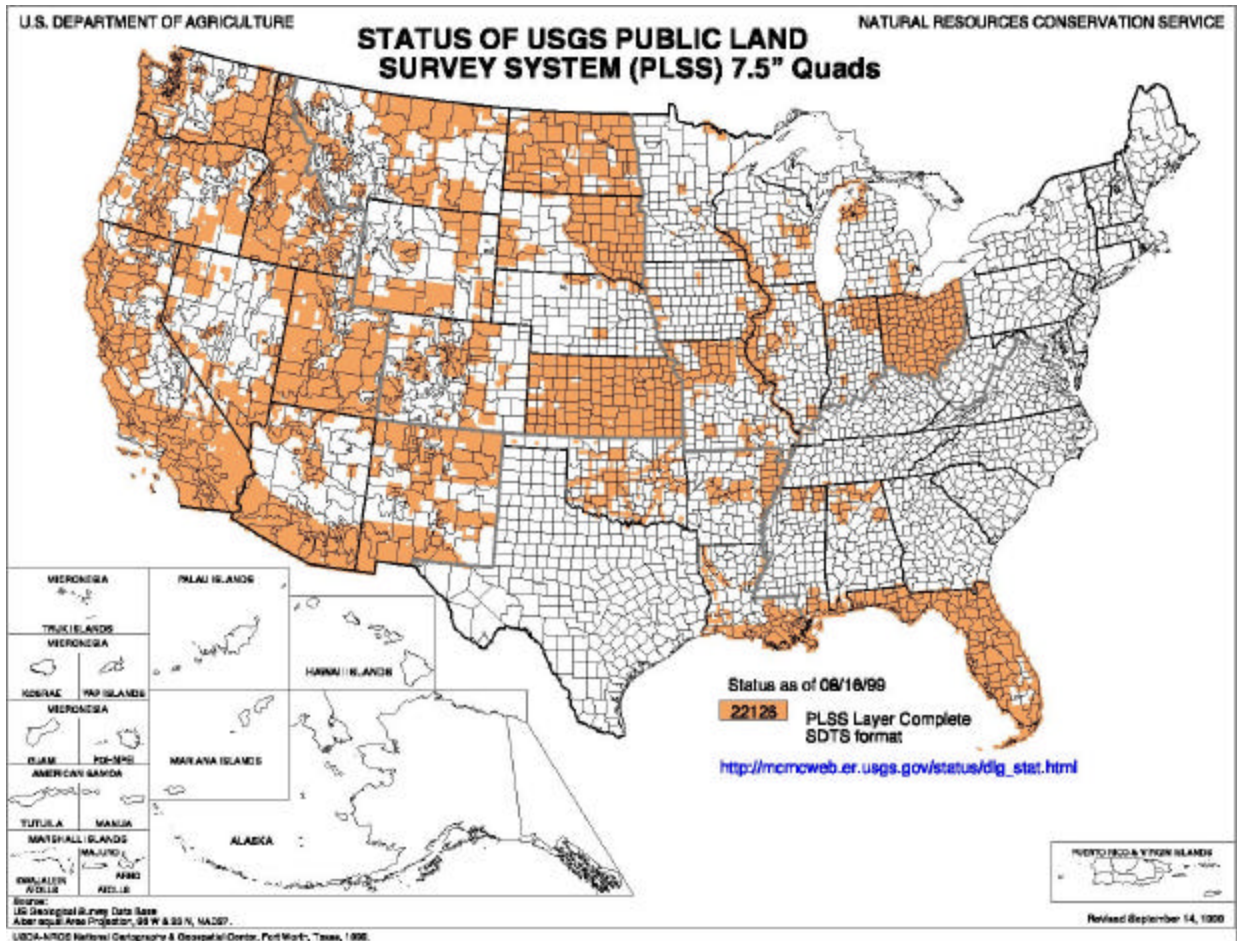


Figure 12 shows the status of the USGS Public Land Survey System (PLSS) data layer that is currently available for acquisition, integration, and delivery to Service Centers. This status information is available in greater detail for display at the state level over the Internet at this URL, http://mcmweb.er.usgs.gov/status/dlg_stat.html. A few states Kansas, North Dakota, Ohio, Utah, and Florida have complete coverage at this time. This data layer can be produced when the 7.5-minute DRGs are prepared.

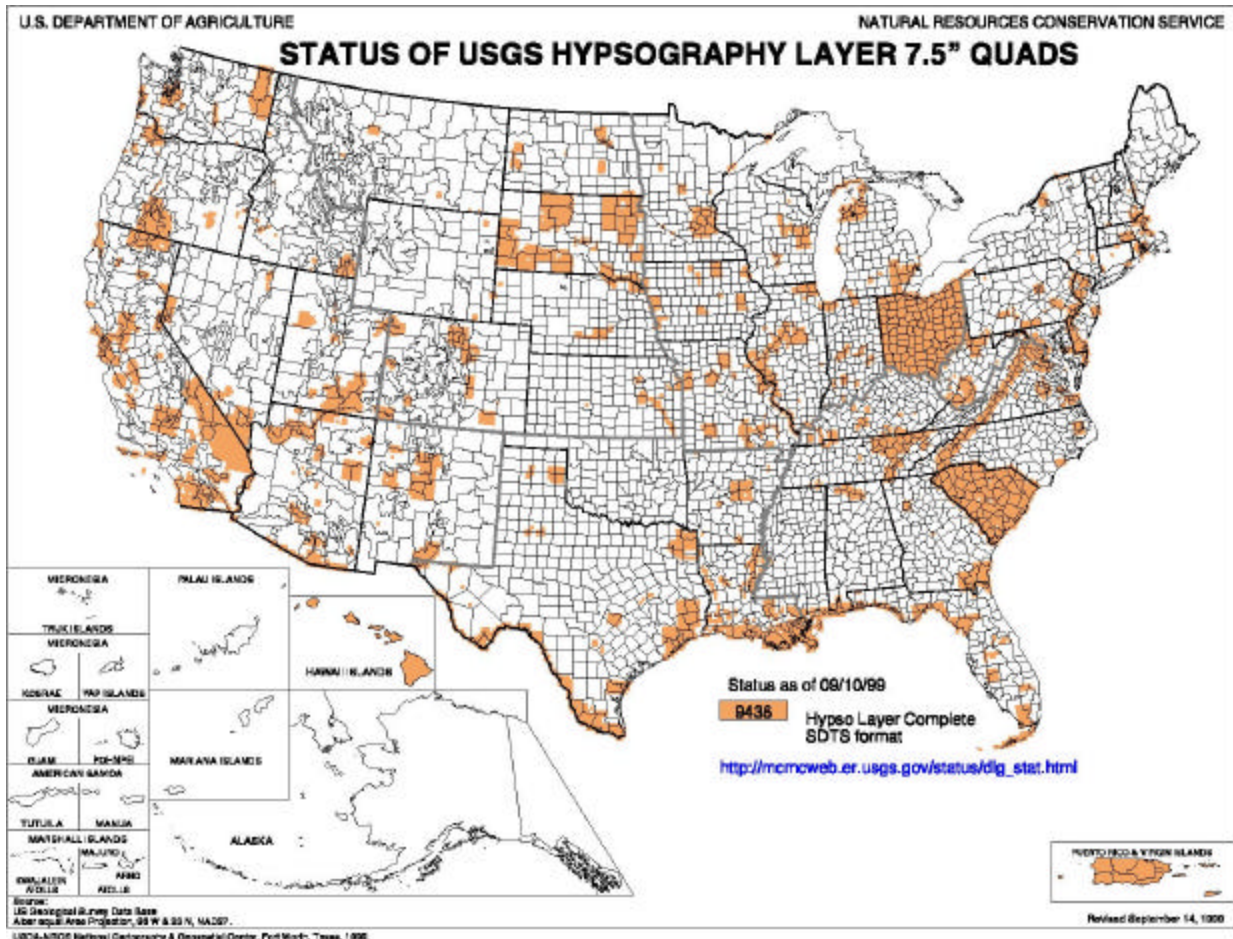


Figure 13 shows the status of the USGS hypsography data layer that is currently available for acquisition, integration, and delivery to Service Centers. This status information is available in greater detail for display at the state level over the Internet at the URL, http://mcmcweb.er.usgs.gov/status/dlg_stat.html. A few states Ohio, Delaware and South Carolina have substantial or have complete coverage at this time. This data layer can be produced when the 7.5-minute DRGs are prepared.

5.3 Theme Descriptions

The table below presents a high level description of each of the critical and non-critical data themes.

Theme	Description
Air Quality	Air quality
Applied Conservation Practices	USDA applied conservation practices maps
Archeology	Archeological sites
Cadastral 24K PLSS	Public Land Survey System captured at 1:24,000
Cadastral PLSS GCDB	Public Land Survey System
Climature/ Precipitation	
• Annual	Annual precipitation by state
• Precipitation monthly mean	Monthly average precipitation by state.

• Temperature monthly mean	Monthly average temperature by state
• PRISM	Climate data from the Oregon Climate Service
Common Land Units	Common Land Units, which include field boundaries digitized from DOQ county mosaics
Cultural/Demographic	
• Census tract boundaries	Census district borders
• Census of Population and Housing	Portray various population and housing statistics
• Economic census	Portray various economic statistics
• Census of Agriculture	Portray various agricultural statistics
Digital Orthophoto County Mosaic (MDOQ)	Digital Ortho Photo Quadrangles that have been mosaicked together for an entire county, compressed to reasonable size.
Digital orthophoto quarter quads (DOQQ)	Rectified, digital photographs, either in B&W or color, that cover one quarter of a USGS quadrangle
Digital orthophoto quadrangles (DOQ)	Rectified, digital photographs, either in B&W or color, that cover a USGS quadrangle
Digital Raster Graphic	Raster version of the 1:24,000 USGS topographic sheets. Includes rivers and roads.
Elevation	Digital Elevation Models, or a raster model of topography.
Endangered Habitat	Endangered species extents
FEMA-Flood hazard maps	Flood hazards from the Federal Emergency Management Agency
GNIS	Geographic Names Information System center points of items, which are shown on 1:24,000 USGS quadrangles. For example, cities, mountains, lakes.
GNIS Cities	Cities and towns from the GNIS layer.
Government Units/ Administrative Bounds Administrative/Political 100K- state,county,city,forests	Various government and other administrative areas
• Managed Area Database	Managed Areas
• Counties 100K	Counties
• Congressional Districts	Congressional Districts
• National parks	National Parks
• Indian Lands	Indian Lands
• National Forests	National Forests
• BLM Lands	Bureau of Land Management Lands
• Military Reservations	Military Reservations
HUC 14-Digit	Hydrological Unit Code (14digit) map at 1:24,000
Hydrography 1:100,000	Surface water features - rivers, streams, lakes
Hydrography 1:24,000	Surface water features - rivers, streams, lakes
Hydrography RF3	Surface water features - rivers, streams, lakes
Index Quads 24K, 100k,250K	USGS quad sheet boundaries
Index Quads 12k	USGS quad sheet boundaries

Index NAPP	Center of aerial photos covered by the National Aerial Photography Program
Land use/Land Cover	National coverage LULC, 1:100,000 or 1:250,000 scale
Land use/Land Cover GAP	Specific state coverage only
Office Information Profile	USDA office locations and associated data
Plants	USDA Plants database
Soils/Prime Farmlands/HEL	Soils from NASIS and SURGO, soil interpretation maps including prime farmlands and highly erodible lands.
Soils/MLRA	Major Land Resource Area polygons
Transportation 100K-Roads,RR,Pipe,Transmission	Transportation at 1:100,000 scale
Transportation Roads	Roads
Transportation Railroads	Railroads
Water Control Infrastructure/Dams	Dams, weirs, etc.
Wetland Determinations	Wetlands determinations from USDA
Wetland - Easements WRP	Wetlands easements from USDA
Wetlands-NWI	Wetlands from the US Fish & Wildlife Service

Table 5-1 High level summary of critical and non-critical data themes

See Appendix A for an example of a detailed theme description. The DRG data set was selected for a detailed description due to the number of production steps involved in preparing this data set for use at the pilot sites. This description follows the integration examples included in Appendix A.

5.4 Business Uses

The following matrix lists the business uses of each theme.

Theme	Uses
Air Quality	
Applied Conservation Practices	Help identify the level of source treatment in counties to determine priorities for directing assistance and funding.
Archeology	
Cadastral 24K PLSS	Help identify customers, for map orientation, and large scale analysis.
Cadastral PLSS GCDB	Help identify customers and for map orientation.
Climate/Precipitation <ul style="list-style-type: none"> Annual 	Used together with crop management practices and yield potentials for analysis of disaster assessment, environmental impact, and risk management programs.
<ul style="list-style-type: none"> Precipitation monthly mean 	“
<ul style="list-style-type: none"> Temperature monthly mean 	“
<ul style="list-style-type: none"> PRISM 	“
Common Land Units	Help administering USDA programs by defining the relationship between customers and the land.
Cultural/Demographic <ul style="list-style-type: none"> Census tract boundaries 	Help identify customers and their attributes, under-served areas, allow comparison of targeted group

	participation in programs that were specifically designed for their benefit.
• Census of Population and Housing	“
• Economic census	“
• Census of Agriculture	“
Digital Orthophoto County Mosaic	Extremely useful backdrop for other digital data, large-scale analysis, and data entry of CLU and other themes.
Digital orthophoto quarter quads (DOQQ)	Extremely useful backdrop for other digital data and for large-scale analysis.
Digital orthophoto quadrangles (DOQ)	Extremely useful backdrop for other digital data and for large-scale analysis.
Digital Raster Graphic	Useful backdrop for other digital data. Provides useful orientation with roads, rivers, rail, and towns.
Elevation	DEMs provide a good small-scale landscape of the Service Center area.
Endangered Habitat	Check proximity of endangered species required for scoring CRP land. CARAA project also requires this data.
FEMA-Flood hazard maps	Provide information for rural housing, facilities, and building site locations, waste, nutrient, and pesticide management.
GNIS	Orientation on the base map. Points are visible on USGS topographic sheets
GNIS Cities	Orientation on the base map. Cities and towns extremely helpful.
Government Units/Administrative Bounds Administrative/Political 100K- state,county,city,forests	Assist with inquiries and making geospatial products for farmers, ranchers, agricultural industry, school districts, or county planners.
• Managed Area Database	“
• Counties 100K	“
• Congressional Districts	“
• National parks	“
• Indian Lands	“
• National Forests	“
• BLM Lands	“
• Military Reservations	“
HUC 14-Digit	“
Hydrography 1:100,000	Orientation and smaller scale analysis.
Hydrography 1:24,000	Orientation and analysis at a larger scale.
Hydrography RF3	Orientation and smaller scale analysis.
Index Quads 24K, 100k,250K	Map orientation
Index Quads 12k	Map orientation
Index NAPP	Determination of photographic coverage
Land use/Land Cover	Conservation Priority Areas and Environmental Quality

	Improvement Program proposals need land cover data set to determine and prioritize natural resource concerns.
Land use/Land Cover GAP	Conservation Priority Areas and Environmental Quality Improvement Program proposals need land cover data set to determine and prioritize natural resource concerns.
Office Information Profile	Information about services provided by nearby offices
Plants	Ranges of plant habitats for restoration and conservation programs
Soils/Prime Farmlands/HEL	Aid in soil and other natural resource conservation
Soils/MLRA	“
Transportation 100K Roads, RR, Pipe, Transmission	Map orientation
Transportation Roads	Map orientation
Transportation Railroads	Map orientation
Water Control Infrastructure/Dams	Estimating potential hazard classification, identifying ownership and agency involvement.
Wetland Determinations	Monitoring of certified wetland determinations and easements
Wetland - Easements WRP	wetland restoration
Wetlands-NWI	General indication of wetland locations for wetlands programs

Table 5-2 Business use of critical and non-critical themes

5.5 Production Costs

The acquisition, integration, and delivery of geospatial data for the Service Center Initiative are scheduled to begin in FY 2000. However, before starting, APFO and NCGC must possess the necessary hardware, software, and human resources needed to complete the project on time. The ramp-up costs estimated to meet this need at both NCGC and APFO are listed below in Table 5.3.

AID Production Facility	Estimated Ramp-up Cost
NCGC – Ft. Worth, Texas	\$277,000
APFO – Salt Lake City, Utah	\$400,000
TOTAL:	\$677,000

Table 5-3 NCGC and APFO estimated ramp-up costs for hardware and software

After the initial ramp-up at APFO and NCGC, the AID processes must be fully engaged, generating countywide data on a consistent basis. The estimated cost to complete the AID production on an average county at each facility is listed below in Table 5.4.

AID Production Facility	Estimated Cost per Average County
--------------------------------	--

NCGC – Ft. Worth, Texas	\$1,500 (1)
APFO – Salt Lake City, Utah	\$4,545 (2)
TOTAL:	\$6,045

Table 5-4 NCGC and APFO estimated AID production costs per average county

Notes: (1) - Integration and delivery of the soils and demographic data themes only.
 (2) – Integration and delivery of the DOQ data only. The CLU data will be produced without any additional integration by APFO.

All of the non-critical data themes will be integrated and delivered by NCGC and state GIS staff as appropriate. The estimated cost per average county to complete this work is \$6,500 each. The funding for these costs is anticipated through the Cooperative Federal Information Partnership (CFIP) and/or other cost sharing partnerships. The CFIP is a coordinated initiative by federal agencies. The effort is intended to support local partnership development for the creation and use of digital geospatial data in support of community planning and resource assessment. At this writing, none of the agency budget initiatives for CFIP funding have been approved. Funding for FY 2000 is highly unlikely. Interest in CFIP continues within USDA, and budget proposals are anticipated for FY 2001 and beyond. However, the initiative is dependent upon agency and department budget levels making long term planning difficult.

The following matrix provides an integration cost comparison by theme.

Theme	Data Producer	Data Publisher	Nominal Scale	Production Costs
Air Quality				Unknown
Applied conservation practices				Unknown
Archeology	Unknown	Unknown	Unknown	Unknown
Cadastral 24K PLSS	USGS	USGS	1:24,000	\$2,160
Cadastral PLSS GCDB	BLM	BLM		
Climate Precipitation annual & monthly mean	Oregon State University	Water and Climate Center - NRCS		\$60
Temperature monthly mean	Oregon State University	Water and Climate Center - NRCS		\$60
Temp & Precipitation Tables				
PRISM	Oregon Climate Service	NCGC-NRCS	1:250,000	
Common land units	APFO, Contractors	USDA	1:4,800	
Cultural/Demographic Census tract boundaries	US Census	Census/ESRI		\$690
Census of Population and Housing				

Economic census				
. Census of Agriculture				
Digital Orthophoto County Mosaic	APFO			\$3,500
Digital orthophoto quarter quads (DOQQ)	USGS, APFO, Contractors	USGS, APFO	1:12,000	Incl.
Digital orthophoto quadrangles (DOQ)	USGS, Contractors	USGS, APFO	1:24,000	Incl.
Digital Raster Graphic	USGS	USGS	1:24,000 to 1:250,000	\$2100
Elevation	USGS	USGS, APFO	1:24,000	\$240
Endangered Habitat				Unknown
FEMA-Flood hazard maps	FEMA	FEMA	1:24,000	Not Determined
GNIS	USGS	USGS	1:24,000	\$30
GNIS Cities				Incl.
Government Units/ Administrative Boundaries				\$60
Administrative/Political 100K- state,county,city,forests	USGS	USGS		
Managed Area Database			1:2,000,000	
Counties 100K	USGS-Water	USGS		
Congressional Districts				
National parks				
Indian Lands				
National Forests				
BLM Lands				
Military Reservations				
HUC 14-Digit	USDA-NRCS		1:24,000	\$240
Hydrography 1:100,000	USGS	USGS	1:100,000	\$2100
Hydrography 1:24,000	USGS	USGS	1:24,000	
Hydrography RF3	EPA	EPA	1:12,000	
Index Quads 24K, 100k,250K	USGS	USGS		\$30
Index Quads 12k				\$30
Index NAPP				\$30
Land use/Land Cover	USGS	USGS	1:250,000 1:100,000	\$450
Land use/Land Cover GAP	USGS/BRD/ GAP	UCSB Biogeography	1:100,000	
Land use/Land Cover MRLC	EPA, USFS, USGS, NOAA,	USGS	1:100,000	

	etc.			
Office Information Profile				Unknown
Plants	USDA/NRCS	USDA/NRCS		
Soils/Prime Farmlands/HEL	USDA	USDA	1:24,000 1:12,000	\$810
Soils-MLRA Major Land Resource Areas				
Transportation - 100K Roads,RR,Pipe,Transmission	USGS	USGS	1:100,000	\$2100
Transportation - Roads	USGS	USGS	1:24,000	
Transportation - Railroads	USGS	USGS	1:24,000	
Water Control Infrastructure/Dams	ACE & FEMA	EPA		Not Determined
Wetland - Determinations	USDA-NRCS			Not Determined
Wetland - Easements WRP	USDA-NRCS			Not Determined
Wetland - NWI	U.S.F&WS, National Wetlands Inventory	U.S. Fish & Wildlife Service, NWI	1:24,000, 1:63,360	\$900

Table 5-5 Integration cost comparison by theme

5.6 Production Schedule

This section details the estimated production schedules for the critical and non-critical data sets. Additional information concerning the projected completion and integration of the critical themes and the production schedules for each of the critical themes is included.

5.6.1 Critical Themes

Based on these estimated ramp -up and AID production costs listed above, the AID Production Schedule would start in FY 2000 and be completed in FY 2007. These dates were determined assuming an annual AID budget of approximately \$2 million. The approximate number of counties to be completed each FY is listed below in Table 5.6. If funding levels increase, the production schedule can be accelerated to produce approximately 600 counties per year and be complete by FY 2005. Production funding beyond FY 2000 can be more accurately determined once the production for FY 2000 has been completed. Table 5-6 represents the production of DOQ, CLU, and soils data only.

Fiscal Year (FY)	Approximate Number of Counties to be Completed
2000	350
2001	400
2002	400
2003	400
2004	400
2005	400
2006	400
2007	400

Table 5-6 Approximate number of counties to be completed each fiscal year

Based on the latest funding estimates for the production of the digital orthoimagery, soils, and common land unit data, production of these themes could be completed by FY 2005, as described by the bar chart presented in Figure 14. Due to variations in the production schedules for each data set, the number of counties completed each FY varies according to theme. The number of counties completed by theme in Figure 14 are used to estimate the number of counties completed in Table 5-6. The production schedule in Table 5-6 is dependent on the successful completion of each data set.

PROJECTED COMPLETION OF CRITICAL GIS THEMES*

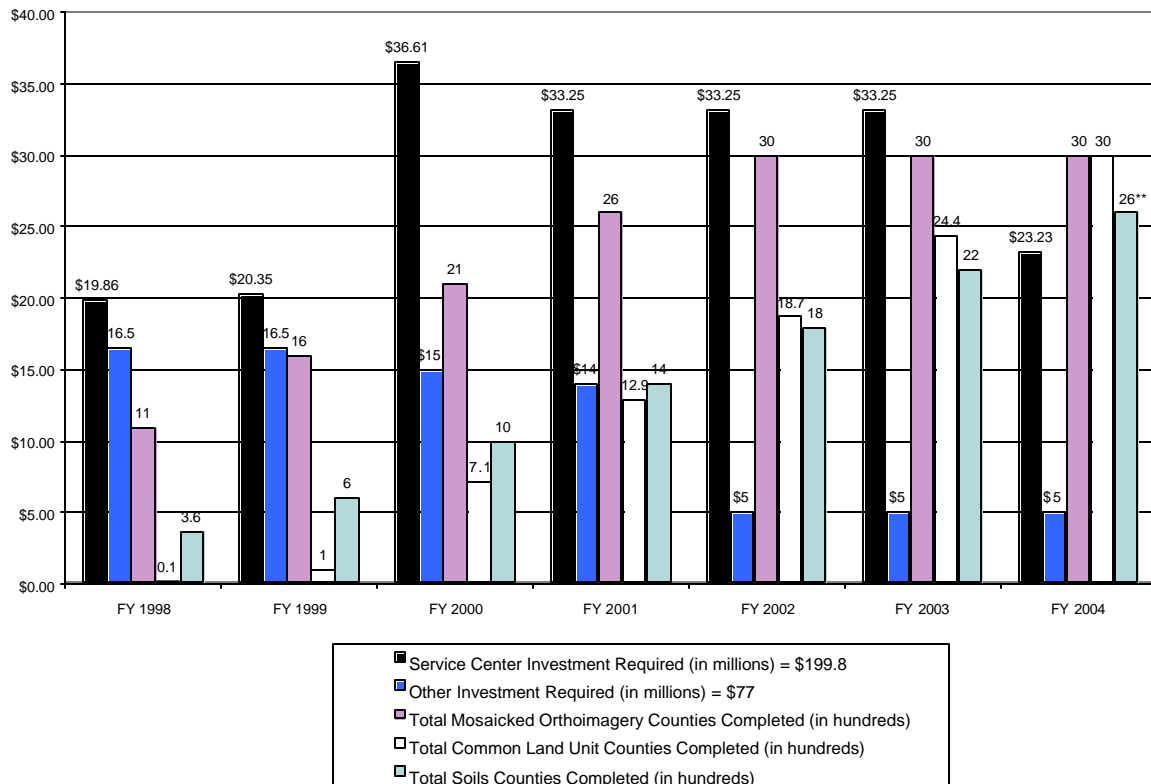


Figure 7 Projected completion of critical GIS themes

*These numbers include the entire U.S. and its territories and affiliated nations.

** Some counties, particularly federal lands, may not have digitized soils.

5.6.2 Non-Critical Themes

Since many of the geospatial data themes required for the implementation of GIS at the USDA field offices are produced by non-USDA agencies, the success of the Service Center Initiative is dependent on close coordination with these other agencies. It is important that the USDA geospatial data production schedule be communicated and synchronized with the data production schedules of these other agencies to ensure all of the required data themes (critical and non-critical) are available for a particular Service Area at the same time.

A matrix of the URLs to information on known production schedules for the non-critical themes is presented below in Table 5.7.

Theme	URL
National Elevation	http://edcwww2.cr.usgs.gov/umap/ned/ned.html
National Hydrography	http://nhd.fgdc.gov
Digital Line Graphs (DLG)	http://mcmweb.er.usgs.gov/status/dlg_stat.html
Digital Raster Graphs (DRG)	http://mcmweb.er.usgs.gov/status/drg_stat.html
Hydrologic Unit	http://imsweb.ftw.nrcs.usda.gov/maps/huc.html

Table 5-7 Estimated production schedules for the non-critical themes

5.6.2.1 Involvement of State GIS Staff with AID of Non-Critical Themes

States and local entities have been aggressively developing digital geospatial data for several years, and often at higher resolution than national data sets. While these data sets may require additional work to be integrated in Service Centers, the data sets are much more useful for Service Center business applications than lower resolution data sets that may be available at the national level. For example, while 1:100,000 hydrography is of very limited use in a Service Center, some states are cost-sharing the development of 1:24,000 hydrography data as part of the USGS-NHD (National Hydrography Dataset). Business applications such as examining stream buffer composition become feasible with such a high-resolution depiction of streams.

The implementation phase of AID will require close involvement of state GIS staff. State staff will acquire, integrate and deliver these state and local data to Service Centers according to the data standards and structures set forth by national AID efforts. They will also coordinate with partners in the development of data that will meet business needs of Service Center agencies.

State GIS staff will also play critical roles in providing support to Service Centers through training, help desk support, cartographic support, and aggregation of spatial themes to meet large-area business needs of GIS such as watershed planning.

Cost, time and staffing considerations need to be made for coverage of these functions at the state level. For example, in many states NRCS has just one staff person handling NRI, SSURGO digitizing and GIS data development and applications. Certainly, additional staff would be needed to cover AID and training functions.

6 AID Assessment

The following section presents an assessment of the geospatial data AID performed for the pilot sites, the prototyping efforts by the Resource Data Gateway team for automating some of the Data AID processes, and the advice given to the team from agencies performing similar tasks during the Best Practices forums. The case studies, problems encountered, and lessons learned are discussed. Recommendations based on these findings are presented in Section 7.

6.1 Acquisition

Geospatial data acquisition is the process of gathering geospatial information that meets the organizational business needs. These data assets can be collected internally (i.e. digital soils surveys) or externally (i.e. digital elevation model). As a result, the scope of the acquisition strategy must be comprehensive to include commercial procurement, partnerships, and cost-sharing as well as internal USDA development.

6.1.1 Problem Statements

Several impediments to successful data acquisition have been identified throughout the pilot implementation, documented through case studies and lessons learned. The problems associated with data acquisition can be summarized into the areas of data availability versus suitability, maintaining quality and consistency, amount of resources required, lack of metadata, and identifying geospatial business needs. A description of how each of these problems impact data acquisition are detailed below:

Data Availability versus Suitability - Various federal, state, and local government agencies and commercial vendors produce geospatial data sets. Each producer generates a data set to meet a set of specific requirements for their business applications. Often times, data sets are duplicates due to a lack of coordination and partnership agreements among agencies that share similar needs. Since these data are often the product of one agency's needs, they often do not meet the needs (for example, it may not have been produced in the correct scale) or cover the geographic extent of another agency (for example, it may not be nationwide).

Quality/Consistency - Maintaining quality and consistency for all data sets delivered to the Service Centers is high priority. Currently, the acquisition of geospatial data includes performing quality checks for each data set to ensure the data matches horizontally at the edges, data sets match vertically when overlain, and attributes are consistent. It is difficult to maintain control over the quality and temporal nature of data sets obtained from external agencies due to the variety of available sources and scales.

There is a lack of consistency in the availability of the geospatial data. Some themes are available nationally others are not. Misalignment of themes is not uncommon. For example, as of September 14, 1999, 673 SSURGO soil survey areas have been certified to meet SSURGO standards. Approximately half of these were compiled on an orthophoto base.

Inconsistencies also exist among the tabular data. For example, the process to acquire the MS Access/NASIS database has changed since the original pilot deployment. The new process is in the planning stages. Another example of inconsistent data availability is within legacy data systems. Only two of the pilot sites has received legacy soils data that is required for CRP eligibility determinations. The process to acquire this data set for all Service Centers in FY2000 has not been determined.

Resources - Data acquisition is the most resource intensive component of any GIS implementation program. Acquiring or building GIS themes is costly in terms of dollars, time, staff, training, and digital storage requirements.

Metadata - Availability of metadata for internally and externally generated data sets is a problem. Metadata is time consuming to produce, and benefits the recipients of the data, not the creators. Data stewards, who are responsible for the collection and maintenance of metadata, are not clearly defined.

Geospatial Business Needs - Currently, there is no clearly defined, finite list of data set needs by business area. Business requirements change frequently, affecting the type of data collected, the scales at which they are collected, and the attributes associated with a particular data set. Business needs cannot be met without standard requirements.

6.1.2 Case Studies

The following case studies clarify the problems encountered during geospatial data acquisition process.

Soils Data Acquisition

Soil data acquired for the Service Centers consist of spatial and tabular data. Spatial data consist of point, line, and polygon soils data derived from Soils Survey Geographic (SSURGO) data. There are two types of tabular data provided to the Service Center with soil information. The first type was delivered to all pilot sites. It is a Microsoft Access database derived from the National Soil Information System (MS Access/NASIS). The MS Access/NASIS database consists of 49 data tables, 10 metadata tables, 12 frequently used queries (reports), and 2 static metadata queries (reports). The second type of tabular soil data is a legacy data set in a Microsoft Excel spreadsheet. These data were provided to only two pilot sites for use in determining CRP eligibility.

Prompt acquisition of spatial data and the resolution of issues regarding them was facilitated by communications with the personnel involved in archival of the SSURGO data. For the pilot sites that were waiting for their SSURGO data, spatial data was integrated and ready to be sent to the I/O Lab the day after it was archived.

The spatial and tabular data are linked by the map unit symbol. These symbols must match exactly for the join to be accomplished. In one of the pilot sites, the spatial data is not certified. The map unit symbol in the spatial soils data table does not match the MS Access/NASIS data and would not join without modification to the spatial data table. Updates or revisions of the map unit symbol in either the spatial or tabular data will require identical changes in the other. This problem occurred at one of the pilot sites. The soil survey has had extensive legend modification in NASIS. More than 10 map unit symbols have been changed since the data was delivered in April of 1999. To use these updated data, the SSURGO data set will have to be updated as well, thereby supporting the need for a refresh management strategy.

Contracting

Data acquisition can be extremely time-consuming. Due to labor costs, outsourcing data collection can almost always be accomplished for less than in-house collection. However, the final quality determination lies within the role of the government. Many agencies try to strike a balance of outsourcing the labor-intensive, repetitive tasks and retaining the knowledge base gained by performing the tasks internally.

Contracting the acquisition of certain geospatial data layers is one alternative to the time consuming and expensive collection of data. Contractors were used to automate CLU data for the nine pilot sites. The cost to automate using contractors and the quality of the

data will be compared to the cost to digitize and the quality produced at the CLU digitizing centers.

Another contracting option is to establish a Cooperative Research and Development Agreement (CRADA) between agencies and contractors or universities. Some agencies, such as NOAA, have established CRADAs with contractors to maintain value-added data. Commercial data vendors are another source of outsourcing. Data vendors often enhance public domain information, packaging it in a plug-and-play format. This could speed the acquisition of quality data to the Service Centers.

Partnering

Agency partnerships are an alternative to acquiring high cost digital data where agency data interests overlap. Cost-sharing partnerships are an essential component of the production of expensive data sets like DOQs. Studies are underway to explore the use of cost-sharing partnerships with USGS for the acquisition of DRG and DOQ for certain areas having incomplete coverage. Additionally, TVA has expressed interest in partnerships to accelerate the production of DOQs. For partnerships to be successful, USDA priorities need to be shared with potential partners in an effort to unite priorities. The ability to collaborate with other agencies seems limited unless department-level or congressional mandates are imposed.

Partnerships for data generation are also taking place at the state level. Many state resource agencies have need for the same type of data as Service Centers, and have funds to assist with data development. One major advantage of state-level partnerships is the willingness of state agencies to pay for higher resolution data for their jurisdictional areas. The disadvantage is that without close coordination, the data may not meet the standards and specifications for the Service Centers.

State Involvement

When NCGC started to look for geospatial data for the pilot sites the first step was to call and E-mail the State NRCS GIS Specialist. This was to ensure that NCGC would not AID geospatial data from a national FGDC clearinghouse if better or more appropriate data was available at the state level. The State GIS Specialist either already possessed the data in question, or knew where to locate it within a state data warehouse. The following examples detail specific cases of state involvement at the pilot sites:

Okeechobee & Glades, FL - David Kriz, NRCS GIS Coordinator provided four CD-ROMs that were developed by the South Florida Water Management District. South Florida has a well-established GIS program and the data was already in ARC/INFO exchange file format. Some of the data obtained were from federal sources, such as USGS. These data were stored in the State Plane coordinate system, and were projected into UTM and converted to shapefiles in order to conform to the rest of the pilot sites.

Miami, KS - Travis Rome, NRCS GIS Specialist provided data from the Kansas Clearinghouse located at Kansas University. Kansas has a very active state GIS program that supports ESRI formats. The Kansas State site is an FGDC clearinghouse node.

Scott, IN – A NRCS employee sent NCGC the Digital Map Finishing layers for hydrography and cultural resources. These data were not in a format supported by the pilot sites and did not have any attributes. However, this data was compiled to fit the DOQ base layer.

Sacramento, CA – The Teale Data Center is a resource for geospatial data. However, since the Teale Data Center charges a fee for their data, the DRGs were the only geospatial data obtained from this source.

Taylor, TX – The Texas Strategic Mapping program is very active and will be a major focal point for AID in Texas. The data development for Taylor was already in development during the pilot effort, so Texas Strategic Mapping's initial involvement was minimal. This program should be monitored for progress and continued involvement with Taylor, TX.

Metadata

Metadata production is far behind data production at most agencies. Currently, there is no comprehensive COTS metadata collection tool available that meets the needs of the user and development communities.

States again may provide assistance with the metadata issues. As a requirement for serving data on the Internet as an FGDC Clearinghouse node, states are developing compliant metadata. In many cases, states have written metadata for national data sets they serve for their state, such as DOQs and DRGs. There are currently about twenty states serving data with FGDC-compliant metadata.

Automation Performance Testing

The Gateway project has tested the possibility and performance of automating data acquisition for well documented, web-available data sets, such as those available on FTP (File Transfer Protocol) sites. It is possible to download geospatial data sets from the web in an automated process. The time and date of the file can be checked against either the file stored on the USDA warehouse or on a database of dates to ascertain if the file on the remote computer has been updated since last obtained by the USDA computer. Whether or not data sets are stored on a USDA machine and updated when the source files are updated or whether or not they are always obtained for a specific request is dependent on the frequency of request for a particular data set. The testing has proved that it is technically possible to automate the acquisition of such data sets. Testing has also shown that the processing involved will not meet the requirements of providing "real-time" data acquisition. It takes a few minutes to an hour to process the data in this way. It would only be practical in a Service Center where the user was notified by email that the acquired and integrated data was now ready on an FTP site or by a CD-ROM delivery service.

6.1.3 Lessons Learned

The following section details the lessons learned according to the Case Studies described in Section 6.1.2.

Soils Data Acquisition

The State Soil Scientist, under the authority of the State Conservationist, is responsible for the soil data (Part 2 Field Office Technical Guide) provided to the Service Centers, some of which is replaced by elements of the MS Access/NASIS soil database. These data are to be certified by the State Soil Scientist prior to its delivery to the Service Center.

Contracting

Given the staff, equipment, and expertise available today, many states cannot successfully integrate GIS data. Sending unorganized raw data to a Service Center will lead to failure in implementing GIS within that Service Center. A good example is enhancing DOQ data by mosaicking to the county level. These MDOQs have proven to be a much more useable product than raw, unmosaicked DOQs. This example supports the need to supply completely integrated data sets to the Service Center.

Taking into account existing staff, lack of equipment fielded at the national centers and the nature of geospatial data collection and metadata creation to be labor intensive, implementation cannot be completed within the five to seven year timeframe. Contracting should be considered in some situations as a option to speed along the data acquisition process as well as a more concerted effort to ramp up and involve state GIS staff.

Partnering

During the Best Practices Review, the Census Bureau described their experiences automating data. The U. S. Census Bureau spent \$300 million to automate 20 million polygons. Quality digitizing can be achieved through contractors at a much more favorable cost than what the Census Bureau spent on data collection. However, there are intangible benefits to digitizing CLU in-house. These include the ability to remain active in the quality assurance and quality control processes and building in-house expertise of employees. In an effort to maintain up-to-date data, other agencies such as USGS continue to struggle with the stress imposed on existing resources.

State Involvement

As the Case Studies demonstrate, the State GIS Specialist played a vital role in AID during the pilot effort. NCGC was able to save time and resources by contacting the State GIS Specialist first, while obtaining the best available data. Additional savings were realized in the time spent on data processing, since many state GIS programs support the ESRI formats in use at the pilot sites. This process should continue to be followed and enhanced during the implementation of GIS at the Service Centers.

Metadata

At a minimum, metadata should comply with the USDA Geospatial Data Set Metadata Standard. The collection of metadata should be incorporated into data development agreements between agencies, even if USDA has to contribute to the cost of metadata collection. Data produced in-house by USDA should incorporate the collection of metadata as part of the data production process.

Automation Performance Testing

Nationwide data sources are often currently available on FTP sites. These data are typically available at no cost and are relatively easy to acquire. Efforts to automate the acquisition of non-USDA data from FTP sites have proven successful for some themes.

Unfortunately, nationwide data sources are frequently not developed at a scale that is useful for most Service Center requirements. The pilot sites are often more knowledgeable concerning the best data available to suit their business needs than are national centers. Careful consideration of state and local data sources are required due to the variation in quantity, quality, and consistency from area to area or state to state. Maintaining standardized naming conventions is difficult when the data standards differs between states.

6.2 Integration

Geospatial data integration are critical processes for providing seamless, plug-and-play geospatial data that meets specific business requirements. The integrated data is derived from a variety of original sources, scales, and business needs. Well-defined geospatial data standards, developed by the Data Management Team, address the format and structure of the geospatial information.

The integration process involves several steps that enhance the display, reproduction, alignment, and cartographic aspects of the data. These “value added” enhancements ensure that the data conforms to nationally consistent, high quality standards. This section describes the problem statements associated with integration, case studies that support the integration process, and the lessons learned.

6.2.1 Problem Statement

In order to meet the challenge of providing geospatial data as part of a turnkey process, integration must address both organizational and technical issues. The challenges encountered during the integration process include providing horizontally and vertically integrated data, ensuring quality control, maintaining standards, production of metadata, technology infrastructure limitations, and ensuring that the data is used appropriately. Agency handbooks, policy manuals, etc. will need to be updated or replaced as necessary to reflect the integration procedures of digital geospatial data.

Horizontal and Vertical Integration - The USDA SCI GIS Strategy implies that the MDOQ layer will be the integration base for all Service Center data. In many cases, the cost to align other data layers to the MDOQ base and with each other far exceeds the marginal benefit from this effort. This is especially true when dealing with highly generalized data typically available from national sources (i.e., USGS, FWS, EPA, etc.). In contrast, detailed data that may be extremely beneficial locally may have wide specification variances from area to area making it difficult to standardize. However, this data may be so localized that the cost/benefit ratio is too low to justify any value adding. Generally, there is a correlation between the cost of improving the accuracy of generalized data and the cost associated with trying to standardize several detailed data sets to one national standard. Both techniques can be costly to achieve, while providing marginal benefits. The goal is to determine the right amount of effort required to produce data that is satisfactory for the business needs.

As new updated digital orthophoto imagery data is developed and becomes available for use, it will be used to replace the image data currently stored at the Service Center or data

warehouse. This new imagery will likely not match-up exactly with the previous imagery data. In these cases, it will not match the vector data digitized over the earlier version of imagery. There is a cost associated with either aligning the new imagery to old imagery or aligning the old vectors to the new imagery. Additionally, there may be quality implications as satellite technology enables the production of more accurate georeferenced orthophoto images while USDA continues to reference an aging orthophoto base.

Quality - As stated above, there is a cost/benefit analysis required for every step of data integration. Quality control can be an expensive part of the process and should be closely examined. Quality assurance/quality control (QA/QC) processes are easier to control for data collected within USDA because it can be built into the data acquisition processes. In order to ensure that data originating outside the agency conforms to USDA data standards, post process quality control processes must be added to this data prior to use within USDA. The result is a much more expensive process, since controls cannot be applied during production. The QA/QC processes should be assessed for the impact each control has on the business. For example, CLU acreage is a crucial piece of information for many Service Center business decisions. So, the benefit of obtaining high quality geometric accuracy is well worth the cost. In contrast, the benefit of producing high quality road networks may be small if they are only used for reference purposes.

Standardization - Standardization of geospatial data has several benefits. Ensuring that data themes are aligned to the same map projection and datum allows vertical and horizontal alignment to be achieved. Additionally, standardization reduces the amount of manipulation required by the user to use the data and ensures that all users are conforming to the same set of requirements. However, there is a cost associated with conforming to standards. Generally, the cost of conforming to a standard rises with more detailed data layers. Like other integration processes, cost benefit analyses should be established to determine when standards conformance validation is sufficient for the business requirements. Additionally, even though data sets independently meet national map accuracy standards, it does not necessarily mean that they will integrate well together as often times the data originate from different sources and are generated at different scales of resolution.

Metadata – The Federal Geographic Data Committee (FGDC) greatly influences the national effort to collect metadata about geospatial data holdings. Although metadata is beginning to become more common, availability of FGDC compliant metadata at all levels is scarce. Data users, rather than the data producers, primarily use metadata. Therefore, metadata for national data sets is potentially useful to a far broader and larger audience than metadata generated for a local data set or other data set which will not be distributed as frequently.

Ease of Use – Data sets, metadata, symbol sets, or legend files provided to pilot sites that are not documented, self-explanatory, or easy to use, are not utilized. Symbol standards are very important for the recognition of features on a map. Lines colored blue are associated with rivers, while railroad track symbology is universally recognizable.

Symbols for soil point and line features are established by national guidelines. These symbols, along with others, were made available to help ease the transition to GIS, but were often not used due to lack of documentation or knowledge of their existence.

Technological Infrastructure - Rapid technological changes continue to occur in the field of geospatial processing (i.e., spatial DBMS, open APIs, Web-enabled GIS, etc.). Many of these technologies have a direct impact on how, where, when, and who performs geospatial data integration. Currently, higher-skilled and relatively expensive software is required to integrate and maintain a reasonable level of geospatial data quality. This has great implications on the technical infrastructure required to support the effort (i.e., telecommunications, licensing, and skill level). Advances in software technology (i.e., on-the-fly topology, automatic edge matching) eliminate some of the "architectural bottle-necks" associated with integration and should be continuously monitored.

Use of Data - The GIS Strategy identifies approximately 20 geospatial layers that would be useful to perform Service Center business. However, the strategy does not specifically identify how each business process uses that data. The user community and BPR projects determine the suitability of data for certain business applications. Since there is a cost associated with integration of any data layer, the BPR projects must take a critical look at what data is necessary to do their business and what data is "nice to have." Much of the data is only germane to a subset of business processes or only needed in relatively localized areas. BPR teams should communicate data requirements to the Data AID Team so that they do not spend resources integrating data that will rarely be utilized.

To date, the AID Team has solicited minimal feedback from the pilot Service Centers. One issue that was clearly communicated was the inability to easily view tabular data linked to the spatial data. For example, it was not easy to display a map legend that contained the soil map unit name for the soil layer.

SSURGO data are produced in geographic (Latitude/Longitude) coordinates. However, SSURGO, as well as all other data sets delivered to the pilot sites, is projected into a common projection system as part of the integration process. This common system is UTM, a zone-based system. Between 250 and 300 county soil survey areas are intersected by a UTM zone boundary. Areas intersected by a UTM zone cannot be displayed simultaneously since the projection parameters across zones vary, unless the data area is extrapolated to all one zone. The result is that users may only view a portion of the data in a county at a time. This situation can cause frustration among users attempting to perform county-based conservation and analysis. Pilot site implementation did not include a county that is intersected by a UTM zone in order to study the impact on the user community and develop solutions. Currently, the GIS software installed at the pilot sites has the ability to re-project vector-based geospatial data into different projection systems at the time of display. This functionality could provide users with more versatility for vector data. However, it is not useful when using raster DOQs as a base because they cannot be re-projected upon display.

Potential for a loss of credibility exists when Service Center personnel and other soil survey users accustomed to hard copy soil maps expect soil map unit boundaries to vertically align with corresponding features on the aerial photograph. There have been a few occurrences where the digital soil boundaries did not match distinct features on the

Digital Orthophoto. Education of Service Center staffs about data limitations due to resolution, scale, and vertical integration is critical.

6.2.2 Case Studies

The following case studies illustrate some of the problems associated with geospatial data integration.

Soils Data Integration

The integration of soil data has uncovered several areas for potential error propagation. One pilot site received a soil data set that contained a datum error. This error created a 150 to 200 foot shift of all the soil polygons covering a particularly intensive agricultural area. The error was detected when the vertical positional accuracy of the soil map unit boundaries was checked against distinct features on the DOQ and DRG. Once detected, this problem was communicated to the data production staff. The spatial data error was corrected prior to delivery to the Service Center.

Another problem, pertaining to the integration of tabular data and geospatial data was discovered during the pilot site implementation. In this case a map unit symbol disagreement between the MS Access/NASIS tabular database and SSURGO derived geospatial data was discovered while integrating these two data sets. This disagreement was communicated to the data stewards and corrected prior to delivery to the pilot site.

Another integration concern is the need to provide proper information and guidance concerning the appropriate use of a data set and any related materials, such as customized symbol sets and metadata. For example, some of the pilot sites received soils data that contained both point and line features, in addition to soil polygons. A symbol set, used to cartographically represent a feature according to specific color and symbol was developed in an effort to standardize feature representation across all pilot sites. This symbol set was based on national guidelines and was provided to all pilot sites during the geospatial data refresh. However, due to lack of documentation or lack of knowledge, these symbol sets have not been incorporated into the daily activities at the pilot sites. This illustrates an example of how enhancements to the data are not realized without proper education.

Pilot Effort

The Pilot effort has proven that it is possible to successfully automate some of the data integration, including re-projection, re-tiling, reformatting, generation of standard Service Center directory structures and file names, and compressing the data into a self-extracting zip file. Currently, automation of most these processes applies only to the vector data format.

The USGS data that is represented on a 7.5-minute quadrangle is stored at the Earth Resources Observation Systems (EROS) Data Center (EDC) in SDTS format. In order to integrate these data for delivery to the pilot sites, these data were downloaded and subsequently processed within ARC/INFO. This processing included steps to edge match data horizontally, produce a seamless coverage, and conform to a consistent

projection, datum and unit of measure. Another process of integration was reconciling the attributes among delineated data sets. For example, attributes among 7.5-minute quadrangles often times do not match. This occurs because each quadrangle is in reality an independent solution for planimetric mapping for a geographic area of interest (7.5-minute by 7.5-minute) at a snapshot in time. A great deal of effort is involved to reconcile attributes in order to create a complete county seamless tile. Once the county tile is delivered to the pilot site, updates to a single quadrangle located within the county causes the whole integration process to be repeated, in order to ensure that the pilot site has the most up-to-date data. The pilot support process illuminated the need to develop alternative solutions to this time consuming process.

Additional attribute reconciliation issues discovered during the development of the hydric soil layer derived from the base soil layer. Although predefined query and report functions exist that facilitate the extraction of interpretations from the MS Access based NASIS database; it does not automatically link to the corresponding geospatial data. One advantage of the NASIS database is the ability to describe the composition of a map unit in detail. However, this feature adds complexity to the geographical representation the tabular data. The capability to develop complex queries from the soil tabular database and render a definitive geographic display appears to be beyond the scope of the typical user at a Service Center. However, reconciling this issue would enable more advanced users to exploit a wealth of information.

Combined Experience

As the use of DOQ data becomes more widespread, USDA has seen an increase in the number of requests for MDOQs from those outside USDA. Agencies, such as USGS have expressed interest in cost sharing with USDA in the development of the MDOQ data sets. The state of Georgia has already created mosaicked data sets by county, using a single UTM zone; similar to the one FSA is sponsoring. However, Georgia did not go to the extra effort of making the county data sets seamless.

Automation Performance Testing

The potential for automating the integration of many of the data sets has been investigated in the Resource Data Gateway project. Not all functions are appropriate for automation. Functions that are subjective or do not have precisely defined rules cannot be automated because human decision-making skills are required. Functions that the Gateway Team tested that lend themselves well to automation include:

Function ¹	Example
Map re-projection	UTM to State Plane
Map re-tiling	Cookie-cutting the data at a county boundary or user-defined shape and combining the data from all Counties into a single state data set
Data packaging	Renaming files to a standard name and directory structure and compressing the files into a zip format

Table 6-1 Examples of integration functions that can be automated

Note 1: The Gateway Team has only tested automated integration on vector data (i.e. not on the DOQ or other images). Work has been started on automated raster data integration by the Gateway Team in concert with MIT.

All three functions listed in Table 6-1 can be accomplished in less than 60 seconds if all data sets are on a local Gateway integration process computer. Remote data sets must first be obtained through download or other digital transfer technique. This adds time the automated process. The actual amount of the additional time will depend on the size of the data set and the rate of transfer. These automated processes support the current pre-

integration steps underway at APFO and NCGC. The functions detailed in Table 6-1 can be automated and simply applied to the pre-integrated data sets. Automation has been found to reduce personnel time, which reduces the cost for some areas of integration. Although some functions can be successfully automated, it may be desirable to perform certain integration steps only when a specific data set is requested, rather than fully integrating all available data sets. Waiting to process data according to requests increases the options available to USDA customers, such as receiving data in State Plane rather than UTM, receiving data for a specific farm rather than an entire county, receiving only the DOQ rather than all data sets, while diminishing the potential to perform unnecessary process steps.

Automated integration will achieve faster response times when processing software is local to the computer holding the geospatial data, and the data are available in a format that can be processed. If automated integration is to occur upon request, the warehousing format of the data and/or the location of the automated integration process software must be considered. These issues must be addressed before decisions concerning the use of automated integration techniques.

6.2.3 Lessons Learned

The following describes the Lessons Learned according to the Case Studies presented in Section 6.2.2.

Soils Data Integration

The pilot effort to integrate soils illustrated the need to develop an interactive interface to associate geospatial data and tabular data. This interface should be intuitive and provide choices to the user that represent appropriate ways to classify and display the tabular data elements. This interface would allow Service Centers to produce soil derivative maps themselves, thus reducing integration time at the national level and reduce disk space requirements at the Service Centers.

A determination of the soil spatial data's positional accuracy relative to other data layers is needed. This is currently a visual review process that cannot be easily automated.

Pilot Effort

DOQs tiled according to Township do not integrate as well as those tiled according to 7.5-minute quadrangle boundaries. However, Service Centers have expressed a need to query and display geospatial data according to the Public Land Survey System (PLSS) Township boundaries. A database index that is linked to the tiles, used to facilitate the display of data by Township boundary, is required for this functionality.

The NASIS and SSURGO map unit symbols need to be compared to insure a proper join of the spatial and tabular data. Where possible, future production of the value added MDOQ and SSURGO must be coordinated to facilitate coincident data availability.

Combined Experience

The national level DOQ naming convention developed by USGS is a viable naming convention for USDA business. USGS confirmed that FSA mosaicking provides

valuable quality assurance for the DOQ data set, since USGS is only able to inspect ten percent of the DOQs they produce.

Automation Performance Testing

The combined efforts of all agencies, the acquisition, integration, and delivery processes must be monitored for technological advances and availability of resources. The AID process must respond to these advances to reduce production time and cost, while maintaining quality control.

6.3 Delivery

The scope of the SCI covers a widely distributed network of field, state and national locations that must have access to geospatial data. This requirement poses a challenge to provide data in a timely manner, given appropriate security considerations. As in the case of the geospatial data integration process, the main objective for the delivery process is to require minimal or zero effort on behalf of the Service Center or any other user. It must be ready to “plug and play” at the Service Center, because the majority of Service Center employees have no GIS expertise.

The media used to deliver geospatial data will evolve. Currently, File Transfer Protocol (FTP), direct download, CD-ROM, or digital tape are options, depending on the amount of data to be delivered. New methods will evolve over the six-year implementation cycle, and the delivery system must remain flexible enough to adapt to these technologies. The movement of large data sets over the air (wireless), or over hard wire (LAN, WAN, Cable TV) is rapidly a developing technology. The delivery system design and architecture must be flexible enough to capitalize on these emerging technologies.

6.3.1 Problem Statements

A number of problems associated with delivery were highlighted during the pilot support process. These problems resulted from an inexperienced user community, insufficient infrastructure, unclear roles and responsibilities, and the lack of a data warehousing architecture. Each of these problems is described below:

Inexperienced Users - Most pilot site users are relatively inexperienced in GIS. They do not necessarily understand the contents of the data sets or their possible uses. Very little metadata exists to help the users with the appropriate use of the data set and any limitations in its use.

Insufficient Infrastructure - The current telecommunications infrastructure is inadequate to support the electronic transfer of large geospatial data sets and files. This limitation dictates that large data sets and files must be delivered on CD-ROM or 4 mm backup tape. The limited communications bandwidth also has a negative impact on moving many smaller geospatial data sets. None of the production facilities now possess a fully automated CD-ROM writing system that could be used to reduce the delivery time.

This problem is compounded the number of requests for geospatial data will increase in the near future, as the quantity and types of geospatial data continue to become available. Requests are anticipated to originate from the national USDA warehouses and the Service Centers. Requests are expected to increase the workload of the delivery system and the staff responsible for delivery.

Unclear Roles and Responsibilities - It is not clear which team or group is responsible for the establishment and management of the geospatial data warehouses or data sources. The roles and responsibilities of the

Data Management Team and the National GIS Coordination Team need to be defined to address these areas.

Data Warehousing Architecture - A key problem in implementing a delivery strategy is the lack of a technical architecture for data warehousing. Various alternatives for the Enterprise data architecture have been suggested and discussed but no final architecture has been defined. This is causing some confusion between data storage and delivery issues.

6.3.2 Case Studies

Pilot Data Delivery and Refresh

During the pilot support process all of the geospatial data developed for the pilot sites were stored at APFO and NCGC. APFO and NCGC packaged these data on CD-ROMs and/or 4mm tapes for delivery to the I/O Lab in Beltsville, MD. In the I/O Lab, the data went through final integration procedures before CD-ROMs and/or 4-mm tapes were cut and delivered to the pilot Service Centers.

Because the BPR project was a pilot, most of the delivery process was not automated. Many lessons were learned along the way to improve the procedures. Unfortunately, funds were not available to support automation improvements.

One alternative for the delivery of geospatial data is to have APFO and NCGC deliver data directly to Service Centers and other users through WWW, FTP, CD-ROM, DVD, or tape. APFO has a sales staff that can facilitate public access to digital (FTP or CD-ROM) and hardcopy data, and NCGC has an operational FTP and CD-ROM generation capacity. During FY 1999, NCGC delivered 2,288 CD-ROMs and 2,998 FTP downloads to customers. NCGC is purchasing a RIMAGE CD-ROM production system to enhance automated production.

FGDC Clearinghouse Node

NCGC is a FGDC clearinghouse node for soils (STATSGO and SSURGO) and climate (PRISM) data. This allows soils and climate data to be searched on the Web, and ordered, by telephone for delivery on CD-ROM. One delivery alternative is to expand NCGC's clearinghouse node capability and develop APFO as an additional FGDC clearinghouse node. Another alternative is to utilize an existing large computer system as a clearinghouse node to serve the data produced for both APFO and NCGC, such as the facilities in Kansas City. Any of these clearinghouse alternatives would allow users to access an existing system for obtaining the geospatial data as needed. The following support and services are being provided by NCGC and should be considered when setting up additional clearinghouse nodes, including clearinghouse nodes set up by state GIS staff:

- Archive copies of soils and National Resource Inventory (NRI) data at an off-site location.
- Data disseminated by tape, CDROM, and FTP on the Internet according to customer needs and desires.
- An "800" number for customers to order data.
- Data stewards assigned to each data sets.
- Technical support for the coordination of graphic layouts, back page contents, and format contracting for reproduction. Services include data formatting, metadata production, and testing for production of CD-ROM.

NCGC is one of six USA gateways in operation. The history of the NCGC clearinghouse node includes:

Date	Accomplishment
5/10/96	Designated NRCS Clearinghouse by Chief Paul Johnson, Chief, NRCS
10/16/98	Established a National Spatial Data Infrastructure (NSDI) Clearinghouse Node and provided searchable index for z39.50 protocol
1/25/99	Established a NSDI Gateway and provided an alternative internet backbone

Table 6-2 NCGC clearinghouse node milestones

Resource Data Gateway Prototype

The Resource Data Gateway prototype is in development to allow any user, Service Center or otherwise, to access, query, download, or request geospatial data. The Gateway is based on the same mechanisms and underlying software as a FGDC clearinghouse node. It has an easy to use interface designed with the non-GIS literate user in mind. Users have the option to order and receive data in the standard SCI format. All data available for a given county can be integrated, packaged, and shipped to a specific Service Center.

Alternative delivery solutions for the Gateway include the option for users to order non-standard data, by selecting a subset of data sets, according to a specific projection and format for an user-defined area not restricted by county boundaries. For example, farmers could request data for a particular field. This alternative supports both UTM and State Plane projection systems, which should satisfy the needs of most state users. Small data sets could be downloaded directly from the Internet, rather than delivered on CD-ROM. The Resource Data Gateway server, once developed, can be operated from APFO, NCGC, and/or any other location assigned to warehousing the data.

However, it is worth noting that bandwidth on existing LWV network will not support this or other download procedures. Tapes and CD-ROM are the only immediate option for data delivery of data sets.

Best Practices

During the Best Practice Forum other federal agencies cautioned the USDA about attempting to take on too much, stressing the limits of available resources at USDA. Currently, USDA does not have the resources to perform large amounts of automated or manual data delivery. The level of resources required to maintain new data delivery schedules and the delivery of updated data has not been determined.

6.3.3 Lessons Learned

This section describes the Lessons Learned from the Case Studies presented in Section 6.3.2.

Pilot Data Delivery and Refresh

The pilot sites are not using most of the non-critical data themes. This may be due to lack of education, experience, or the layers may not be appropriate. To avoid spending resources on data that are not utilized, the delivery of the geospatial must be linked to the nominal business needs of the Service Centers.

Geospatial data should be delivered in a consistent directory and file structure to facilitate support and management. Efforts to minimize the burden of loading GIS data into the Service Center server should be made. The delivery of data as a self-extracting compressed file or providing uncompressed data that conforms to a standardized directory structure should help the users. Additionally, consistent data update and maintenance policies must be documented to ensure standardization.

Whether to deliver data in a State Plane versus UTM projection will impact potential partnerships with states not interested in UTM data. Therefore, it is important to define the delivery specifications for each of the geospatial data themes.

Additionally, some geospatial data themes are so dynamic that by the time they are processed and delivered, they are already out-of-date. Other geospatial data are more static, and require little or no maintenance over extended periods of time. This situation is complicated because some data sets will be enhanced at the Service Center, resulting in the local version being more up-to-date than the version stored at the state or national level. The process to “upgrade” the national level data to the local standards has not been developed. Processes also need to be developed to aggregate data developed locally into state and national themes to meet various business requirements.

FGDC Clearinghouse Node

Clearinghouse nodes provide a low cost delivery alternative because no specialized software development is required. A clearinghouse node makes data available to both internal (USDA) and external customers and supports compliance with Executive Order 12096 for metadata production. On the down side, there is an assumption that users of clearinghouse nodes are GIS literate. Also, the standard clearinghouse node interface does not offer an easy order or delivery interface. These features are requirements for Service Center data delivery.

The SCI must provide access to digital spatial data through metadata within the framework of NSDI and the FGDC Clearinghouse as mandated by Executive Order (EO) 12906. This EO can be found at <http://www.npr.gov/library/direct/orders/20fa.html>.

Partnerships for data delivery should be established and maintained between the Open GIS Organization and FGDC.

Resource Data Gateway Prototype

Developing a customized, single gateway to geospatial data provides an automated method for data delivery. The advantages include effortless delivery, the ability to process internal and external orders, lower resource costs, the ability to automate many of the data acquisition and integration functions, and the utilization of metadata in compliance with Executive Order 12096. The disadvantage to this delivery solution is that the initial cost to develop a Gateway will exceed the cost to establish a clearinghouse node.

Automated download and delivery solutions like the Gateway and clearinghouse nodes require less resources to deliver data than manual techniques, such as CD-ROM delivery.

Best Practices

USDA should not attempt to provide more than they can realistically support during the initial stages of development. Participants in the Best Practices reviews indicated that development should occur as needed, rather than attempting to deliver too much too soon.

7 Recommendations

An important part of the Geospatial Data AID National Implementation Strategy Plan is the formulate and present Geospatial Data AID Team recommendations. Throughout this plan options are described according to the AID process they address, how they evolved, and how they will benefit future implementation. Many of the proposed recommendations contain underlying recommendations that support or contribute to the implementation of the main recommendation. A comprehensive set of recommendations are described in detail and presented in this section.

1. Fund APFO and NCGC at \$2 million consistent with the President's FY 2000 budget to acquire, integrate and deliver Mosaicked Digital Orthophoto Quadrangles (MDOQ), Common Land Unit (CLU), and soil data for approximately 350 counties. Additional funds in future years will be based on production experience and process efficiencies.
 - 1.1. The \$2 million includes an estimated production cost of \$1.323 million and \$677,000 for hardware and software infrastructure.
 - 1.2. States are now identifying 600 priority counties for FY 2000. All critical themes may not be available for all 600 sites. The 600 priority counties should only be analyzed against the 350 that can be produced in FY 2000.
 - 1.3. Existing non-critical data layers should only be delivered to additional counties as requested and resources allow, keeping in mind that data themes must add value and meet core business needs.
 - 1.4. The cost and scheduling to update, maintain, and refresh data need to be reconciled.
2. Establish a National GIS Coordination Team to implement the recommendations in this strategy and guide the GIS implementation. This Team should be composed of appropriate Business Process Re-engineering (BPR) team representatives, data stewards, existing AID representatives from APFO and NCGC, national headquarters GIS leads, state GIS specialists, and a representative of the Support Services Bureau (SSB) Data Management Team. The Deputy Chief of Soils Survey and Resource Assessment of NRCS and Deputy Administrator for Farm Programs for FSA should be executive sponsors. The Team's responsibilities would include:
 - 2.1. Oversee the utilization of existing resources and partnerships, such as the APFO-NCGC charter.
 - 2.2. Take an active role, along with the partners charged with leading the development effort, to establish funding, and priorities as business areas identify the need for a data set that USDA will not develop independently. Partners in this effort include federal agencies (USGS, BLM, etc.), state and local agencies (DOT, county government, etc.), and other organizations such as universities.
 - 2.3. Coordinate national priorities for MDOQ, CLU, and soil data to implement county GIS.
 - 2.4. Provide input to the design of a system architecture, particularly the geospatial data structure and the subsequent management and update of warehouse and the Service Center data.
 - 2.5. Provide input and coordination for developing data standards.
 - 2.6. Establish partnerships with federal, state, and local agencies and groups at the national level to support geospatial activities.

- 2.7. Ensure that state partnering efforts conform to standards developed by the Data AID and Data Management Teams.
 - 2.8. Provide business and technical input and support to the SSB Information Technology (IT) Development Centers on tools and systems supporting data acquisition, integration and delivery.
 - 2.9. Oversee production priorities. Develop a single priority list that synchronizes the production of the four critical themes. These priorities must be developed with and agreed to by RD, FSA, NRCS, and the national FAC. A policy to manage (modify) the priorities will be developed.
 - 2.10. The National GIS Coordination Team should develop, on a theme by theme basis, an operational plan that includes the identification of geospatial data sponsors and stewards, definition of their respective roles and responsibilities, description of the data model, and change control process for that data theme. Key to this plan include design of the system architecture, including warehousing and delivery structure, the subsequent management and update of data between the data warehouse and the Service Centers. In order for this plan to be implemented, it is recommended that NCGC, APFO, and the Resource Data Gateway team continue to work together to improve and automate the geospatial data AID process.
 - 2.11. Clearly define, understand, and execute roles and responsibilities under SSB to ensure that there is clear accountability of the implementation of the acquisition, integration, and delivery of geospatial data to the Service Centers. Drawn upon from existing roles and responsibilities defined for NCGC and APFO, these roles and responsibilities will expand to include how the three agencies will work together under the umbrella of SSB. Management of the geospatial data must be a partnership between the owners of the data, NCGC, APFO, and state GIS specialists under the framework of the SSB. These roles and responsibilities will be an integral component of the implementation plan.
3. Continue to fund research and development efforts in data acquisition, integration, and delivery.
 - 3.1. Continue research and development efforts at Fort Collins and other SSB Development Centers to evaluate, develop, and deploy tools and systems which facilitate delivery, such as the Resource Data Gateway (\$588K).
 - 3.2. Continue research and development efforts at APFO (\$450K), NCGC (\$210K) in the areas of production automation, process improvement, data maintenance and delivery.
 - 3.3. Provide APFO and NCGC with the funding necessary to keep pace with the technology and resources required to adequately and efficiently integrate the geospatial data according to the implementation schedule.
 - 3.4. Provide resources to states to develop collateral AID efforts in support of those at APFO and NCGC.
4. Deliver geospatial data to Service Center offices as a turnkey process that minimizes the data management requirements at the local level.
 - 4.1. Delivery of geospatial data should be organized, documented, publicized, and made available to customers through a single, virtual gateway access system. This system accommodates NSDI clearinghouses, state partnerships, and Service Center data management.
 - 4.2. Establish separate delivery solutions for servicing normal business at the Service Center, where data can be delivered on CD and served locally, versus delivery to producers from any Service Center. These delivery systems should be adaptable to any technical architecture selected (centralized, decentralized, etc.).
 - 4.3. The telecommunications infrastructure should provide the necessary bandwidth to support the timely delivery of geospatial data and a foundation to support Web geospatial processes in the future.
 - 4.4. Clearly define data access privileges for all potential users and establish guidelines for the reproduction and redistribution of geospatial data.

- 4.5. Access to geospatial data that in the public domain should be made readily available to the general public. Establish password-protected access to geospatial data sets that are not available to the general public due to privacy issues. Data stewards will establish the access requirements for each data set.
5. Monitor geospatial data requirements as business areas identify the data sets that are required to support their business needs.
 - 5.1. USDA data stewards should determine the cost/benefit of acquiring public domain versus commercial geospatial data to meet business requirements.
 - 5.2. Justify business applications that can be accomplished in situations where the four critical themes are not aligned for synchronized delivery.
 - 5.3. Guidelines shall be established to assess the cost/benefit of a new data theme in relationship to acquisition cost and quality assessment costs.
6. Enforce adherence to geospatial data standards, as recognized by the Data Management Team, for development and documentation of geospatial data sets.
 - 6.1. All geospatial data sets should adhere to the Geospatial Data Standard, developed by the Data Management Team. Currently, the data standard outlines processes that are performed during the integration process, such as re-projection to a common projection, coordinate system, and datum, and tiling of data sets along edges. This standardization should continue to be handled during the integration process until users obtain the necessary software tools to easily handle the variations themselves.
 - 6.2. An USDA cartographic symbol set should be developed and used as the national standard.
 - 6.3. Metadata for each geospatial data theme should be collected, organized, and maintained by the data steward for each theme. Metadata should be modified during the integration process to reflect any value-added processes that were applied to the data set.
 - 6.4. Metadata should be delivered along with the geospatial data. This will help users understand the geospatial data, associated tabular data, proper uses of the data, and any restrictions or caveats associated with the data.
 - 6.5. When capturing or procuring new data through the use of a contractor, the collection of metadata, according to the Geospatial Data Set Metadata, should be written into the contract as a delivery requirement. Additionally, the collection of metadata should be incorporated into data development agreements between agencies even if USDA has to contribute to the cost of metadata collection. Data produced in-house by USDA should also incorporate the collection of metadata during the data production process.
7. Educate USDA users about geospatial data and their use limitations due to scale, resolution, map projection, and geometric and temporal mismatches. The GIS Training Team's materials should thoroughly address these issues.
 - 7.1. Service Centers should be authorized to collect geospatial data on an ad hoc basis when there is an immediate need for the data. Guidance and procedures for this will need to be developed and made available.
 - 7.2. Ensure that Agency operational handbooks related to imagery are updated with respect to GIS.
8. Clearly define data access and data reproduction privileges for all of the geospatial data.
9. Monitor acquisition, integration, and delivery processes and respond to technological advances and availability of resources.
 - 9.1. Work with commercial vendors to increase the functionality of COTS software that address specific USDA business needs.
 - 9.2. MDOQs should be the base for developing data that does not already exist and that falls within the domain of USDA responsibility. The exception is when more accurate global positioning system (GPS) data is available.

- 9.3. Establish a quality control and review process to ensure that data themes are accurate and business needs are satisfied.
 - 9.4. Geospatial data maintenance policies and procedures should be documented and provided to the implementation team. This guidance will be used by the National GIS Coordination Team to establish both upstream (local to national) and downstream (national to local) data delivery solutions.
10. Involve state GIS staff in the SCI program, particularly with AID activities. State GIS staff should be provided the tools needed to assist APFO and NCGC with delivery of nationally-consistent, integrated geospatial data to Service Centers, as well as standards and specifications for supplementing higher-resolution state and local data.
- 10.1. Actively solicit input from state staff on data availability to meet business needs.
 - 10.2. Encourage partnerships at state and local level for data development.
 - 10.3. Augment state staff engaged in GIS support activities.

8 Risk Assessment and Management Implications

The implementation of new technology and subsequent re-tooling of business processes generates tremendous challenges for most organizations. Historically, national GIS implementations have not been fully successful for a multitude of reasons. Lack of long-term commitment as organizational priorities change and the absence of resources resulted in projects being abandoned.

The success of GIS, as with any digital technology, relies on several factors. A shared vision from field staff that geospatial tools can assist and support them in daily activities is needed to implement reengineered processes. Top management must be committed to ensure cross-agency cooperation. Management needs to continually work with the Office of Management and Budget and Congress to ensure critical components are funded.

Lacking intuitive, easy to use software and the computer processing capacity to analyze the data, staff will view the tools as inefficient and the technology a hindrance rather than a tool. Training and proper system implementation is needed to reduce the learning curve. Insufficient, inadequate or lack of data in meeting agency business requirements will hamper reengineered business processes and eventual project success. In the absence of consistent funding for people, data development/delivery, hardware, software and training, the Secretary's vision of one stop shopping will not be realized.

Agencies need to support the recommendations in this plan if we are to realize the full benefit of the historical investments made in geospatial data development. Simple, easy access to GIS software will offer value to field employees and equip them to do more as staffing levels decrease continually. Millions of dollars have been spent in the last ten years developing, acquiring and integrating geospatial data for national, regional and local applications. An integrated data management methodology is needed to avoid a duplication of efforts and data, and to ensure the needs of all agencies are being satisfied.

In accepting the recommendations put forth in this plan, the agencies are accepting a variety of risks, which if adequately managed will be outweighed by the benefits. Agencies are committing resources to the continued development and maintenance of geospatial data and associated software tools, however at an accelerated rate. This commitment will need to continue for at least five years. It is envisioned that benefits will be seen through long-term cost savings resulting from shared data resources, new hardware/software, and re-engineered business process. Corporate numbers of conservation practices and program applications will be easier to derive and maintain, while partners and cooperators will have quicker, greater access to data. National, regional and wide area resource analysis will be common for resource assessment and long term agency strategic planning. Staff will have the capacity to address resource issues in a timely and efficient manner using many of the same tools used by their customers.

Potential hidden costs include, research and development needs to analyze, store, retrieve and deliver data. Substantial resources are needed for software maintenance, hardware upgrades, data theme procurement, network upgrades, data delivery medium, high end output devices, help desk support, IT support requirements, and network costs as data downloads over slow networks become more common. Independent prioritization of agency data set development will be done by a separate coordination body. Agencies will need to coordinate their business needs with those of other agencies to save money and gain efficiencies. Since at this writing, the roles and responsibilities of the Support Services Bureau are not yet clearly defined, the impact of this new organizational structure is not known. However, it is assumed they will play a large role in data management.

As described in the USDA GIS Service Strategy document, over 34% or 168 million dollars of BPR cost savings can be attributed to GIS implementation. These savings will not be attained if resources, whether people, data, or money are not continually made available. The annual, incremental funding of SCI is a limiting factor and leads to uncertainty on whether the project goals will be reached.